

OM of: US-09-510-332-1 to: EST:\* out\_format: pfs  
Date: Mar 15, 2001 1:28 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

# Command line parameters:

-MODEL=frank\_p2n\_model -DEV=xlp  
-O=/cgn1\_1/OSPO.spool/US09510332/runat\_06032001\_093637\_13621/app\_query.fasta\_1.359  
-DB=EST -QFMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=6.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELXOP=6.000  
-DELXEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ps -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09510332@cgn1\_1\_2364 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPHY -WAIT -THREADS=1

# Search information block:

Query: US-09-510-332-1  
Query Length: 299  
Database: EST:\*  
Database sequences: 7991742  
Database length: 791223438  
Search time (sec): 491.860000

# score\_list:

Sequence	Strd Orig	ZScore	EScore Len	Documentation
gb_gss4:AO308694	+ 1165.00	2183.32	1.3e-112	742   AO308694 CITBI-EI-2530B8.TF CIT
gb_gss4:AO316999	+ 1106.00	2073.29	1.8e-106	650   AO316999 CITBI-EI-2530B7.TF CIT
gb_gss4:AO277039	+ 455.00	851.56	2.0e-38	310   AO277039 CITBI-EI-2516A7.TF CIT
gb_est7:AA416581	+ 406.00	754.77	5.0e-33	463   AA416581 zu05e04.s1 Soares.test
gb_gss4:AO616191	+ 183.00	359.99	5.3e-11	498   AO616191 HS.5148.A2.F05.SP6E.RH
gb_est28:AL037695	+ 183.00	331.68	1.8e-09	554   AL037695 DKFPP5640672.F1.564.0
gb_gss20:AZ396348	+ 170.50	305.44	6.3e-08	710   AZ396348 IM0160C2IR Mouse 10kb
gb_gss21:AZ416304	+ 169.50	304.39	6.1e-08	656   AZ416304 IM0191D04R Mouse 10kb
gb_gss15:AO711250	+ 157.50	283.83	8.5e-07	538   AO711250 HS.5349.B2.H04.T7A.RQC
gb_gss15:AO711250	+ 150.00	267.76	6.7e-06	644   AZ078264 RPCI-23-399P19.TV.RPCI
gb_est6:BE034300	+ 148.00	257.00	2.4e-05	1176   BE034300 ME01F08 ME Mesembryat
gb_gss23:BI17827	+ 147.50	266.46	7.9e-06	466   BI17827 347119.TPB CIT978SAL.HG
gb_est75:BE734359	+ 141.50	242.45	0.0002	1544   BE734359 601565606F1 NIH.MGC.2
gb_est79:BE037592	+ 135.50	231.40	0.0007	1502   BE037592 601461160F1 NIH.MGC.4
gb_est68:BE250598	+ 135.50	230.40	0.0008	1652   BE250598 600943379F1 NIH.MGC.1
gb_est77:BE872484	+ 135.00	229.22	0.0006	993   BE872484 601447162F1 NIH.MGC.6
gb_est38:AV706129	+ 134.00	232.95	0.0006	993   AV706129 AV706129 ADB Homo sap1
gb_est75:BE781035	+ 134.00	229.48	0.0009	1378   BE781035 601469204F1 NIH.MGC.6
gb_est68:BE215509	+ 133.00	228.02	0.0011	1324   BE215509 HV.CEBD006P23f Horden
gb_est70:BE383624	+ 132.50	221.99	0.0007	832   BE383624 601287177F1 NIH.MGC.19
gb_est78:BE383624	+ 132.50	224.91	0.0016	1625   BE383624 601565058F1 NIH.MGC.6
gb_est76:BE662876	+ 131.50	224.44	0.0017	1420   BE662876 601567644F1 NIH.MGC.6
gb_est70:BE371874	+ 130.50	220.97	0.0027	1649   BE371874 601217970F1 NCI.CGAP
gb_est70:BE371874	+ 130.50	220.96	0.0027	1651   BE748722 601571776F1 NIH.MGC.5
gb_est75:BE748722	+ 130.50	222.62	0.0022	1291   AM014493 EST345797 Normalized
gb_est50:AW914493	+ 129.50	224.57	0.0017	982   BE388502 601287177F1 NIH.MGC.14
gb_est77:BE874969	+ 129.00	219.23	0.0034	1487   BE874969 601487528F1 NIH.MGC.9
gb_est77:BE874969	+ 128.50	221.38	0.0026	1111   BE300630 601673549F1 NIH.MGC.2
gb_est77:BE875629	+ 128.00	221.97	0.0035	1395   BE875629 601235518F1 NCI.CGAP
gb_est72:BE832096	+ 128.00	222.36	0.0021	817   BE832096 601235518F1 NCI.CGAP
gb_gss17:AZ202787	+ 128.00	222.37	0.0022	925   AZ202787 SP_0081.A2.F02.SP6E.S4
gb_est66:BE036798	+ 128.00	221.53	0.0027	1050   BE036798 MP06801 MP Mesembryat
gb_est70:BE396460	+ 128.00	216.07	0.0037	1329   BE396460 601497156F1 NIH.MGC.7
gb_est77:BE906109	+ 127.50	222.47	0.0022	838   AZ209601 SP_0138.A1.C01.SP6E.S4
gb_gss17:AZ209601	+ 127.50	220.10	0.0034	1049   BE782836 601472342F1 NIH.MGC.6
gb_est75:BE782836	+ 127.50	219.20	0.0034	1142   BE782836 MP08611 MP Mesembryat
gb_est66:BE036906	+ 127.50	219.14	0.0034	1148   BE781044 601565055F1 NIH.MGC.7
gb_est75:BE791044	+ 127.50	216.38	0.0048	1490   BE966571 601602023F1 NIH.MGC.7
gb_est71:BE545455	+ 127.50	215.98	0.0051	1548   BE545455 HVSMBH0094K19f Horden

gb\_est68:BE214090 + 127.50 215.36 0.0055 1640 | BE214090 HV.CEBD002D06f Hor  
gb\_gss17:AZ196311 - 127.00 220.73 0.0028 904 | AZ196311 SP.1031.B2.C06.T7A  
gb\_est78:BE960997 - 127.00 218.89 0.0035 1075 | BE960997 601648553P1 NIH.MG  
gb\_est66:BE040951 + 127.00 217.34 0.0043 1245 | BE040951 OF15A01 OF Oryza s

# seq\_name: gb\_gss4:AO308694

seq\_documentation\_block: 742 bp DNA GSS 22-DEC-1998  
LOCUS AO308694  
DEFINITION CITBI-EI-2530B8.TF CITBI-EI Homo sapiens genomic clone 2530B8, DNA  
sequence.

ACCESSION AO308694  
KEYWORDS GSS.  
SOURCE GI:4040728

# ORGANISM

human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 742)

REFERENCE  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
Venter,J.C.

TITLE  
Map Building  
Unpublished (1998)

COMMENT  
Contact: Shaying Zhao, William Nierman, Mark Adams  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel.: 301 838 0200  
Fax: 301 838 0208  
Email: hbbettgr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21  
Class: BAC ends.

# FEATURES

source 1..742  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2530B8"  
/clone\_1ib="CITBI-EI"  
/sex="male"  
/cell\_type="sperm"  
/note="Vector: pBE03AC11, site\_1: EcoRI; site\_2: EcoRI;  
Caltech Human BAC Library D"

BASE COUNT 176 a 175 c 144 g 247 t  
ORIGIN

alignment\_scores:  
Quality: 1165.00 Length: 226  
Ratio: 5.155 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-510-332-1 x AO308694 ..

Align seg 1/1 to: AO308694 from: 1 to: 742

74 giupheilemetysserAlaascysAlaileuleupheileasrgi 90  
|||||  
|||||

2 GAATTCATCATGCTTGCATATGTCATATCTTATTTATTAAGA 51  
|||||

90 uleugluetrpluAlaThrtrleuglyValphetrycysAlaIysv 107  
|||||

52 ATTGAATTTGGCTTGCACATGCGCTGCGCTTTCTATTGTGCAAG 101  
|||||

107 AlAlaserValargHisproleupheiletrpleuIysmetargIleser 123  
|||||

102 TTGGCAGCGTCGTCACCATCTTCATGCTTGAAGATGAGATATCC 151  
|||||

```

124 LysLeuValProTrpMetIleLeuGlySerLeuLeuTyValSerMetI1 140
152 AGCGTGCCTCCATGATGATCTCGGGGTCTCGTATATGATCATATCAT 201
140 eCyValAlaPheHisSerLysTyValAlaGlyPheMetValProTyPheLeuA 157
202 TTGTGTTTCCATACCAATATGACAGGGTTTATGCTCCCATCTCTCCAA 251
157 rGlyPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeuA 173
252 GGAAATTTTTCCTCCCAAAATGCCAACATTCAAAAAGAAATACACTGCT 301
174 IleGlnIlePheSerPheValAlaGluPheSerValProLeuLeuIlePh 190
302 ATACGATTTTCTCTTTGTTGCTGAGATTCTCAGTGCATTCCTTACTCT 351
190 eLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArg 207
352 CCTTTTTCCTTTTGGCTTGTGATTTTCTCTGAGGAGGACACCCGGC 401
207 ImeLysArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaPro 223
402 AAATGAGAAACACAGTGGCCGCGACAGAGGTCTCTGGCAGGGGTGCACC 451
224 IleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyPheSerH1 240
452 ATCAGCGGTGCTGTCTATCTCTCTCTCTGATCTCTGATCTCTACTCTCC 501
240 sCyMetIleLysValPheLeuSerSerLeuLysPheHisIleLeuArgP 257
502 CTGCATGATTAAGTTTTCCTCTCTCTCTCTCAAGTTTTCACATCAGAAGT 551
257 heLePheLeuPhePheIleLeuValIleGlyIleTyProSerGlyHis 273
552 TCATCTTTCTGTTCTTCATCTCTGATGTGTATATACCTCTGCGAC 601
274 SerLeuIleLeuIleLeuGlyAsnProLysLeuGlnAsnAlaLysLys 290
602 TCTCTCATCTTAATTTTAGGAAATCTCAAAATGTAAACAAATGCAAAAA 651
290 sPheLeuLeuHisSerLysCysGln 299
652 GTTCCCTCTCCACAGTAACTGTCTGAG 679

seq_name: gb_gss4: A0316999

seq_documentation_block:
LOCUS A0316999 650 bp DNA GSS 22-DEC-1998
DEFINITION CITR1-E1-2530B7.TF CITR1-E1 Homo sapiens genomic clone 2530B7, DNA
sequence.
ACCESSION A0316999
VERSION A0316999.1 GI:4040265
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venier, J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:

```

```

http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1fb="CITR1-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11, Site_1: EcoRI, Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 149 a 152 c 125 g 224 t
ORIGIN

alignment_scores:
Quality: 1106.00 Length: 216
Ratio: 5.120 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-510-332-1 x A0316999
Align seg 1/1 to: A0316999 from: 1 to: 650

74 GluPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnG1 90
2 GAATATCATCATGTGTTCTGCGAATGTGCAATCTCTTATTTATAATGCA 51
90 uLeuGluLeuTrpLeuAlaThrTrpLeuGlyValPheTyrcysAlaLysV 107
52 ATTGGAATCTTGCTTGCCACATGCTCGCGCTTTCATTTGTCGCAAG 101
107 aLAsSerValArgHisProLeuPheIleTrpLeuLysMetArgIleSer 123
102 TTGCCAGCTCCGTCACCCATCTTCATCTGTTGAGATGAGATATTC 151
124 LysLeuValProTrpMetIleLeuGlySerLeuLeuTyValSerMetI1 140
152 AGCGTGCCTCCATGATGATTCCTGGGCTCTCGTATATGATCATGAT 201
140 eCyValAlaPheHisSerLysTyValAlaGlyPheMetValProTyPheLeuA 157
202 TTGTGTTTCCATACCAATATGACAGGGTTTATGCTCCCATCTCTCAA 251
157 rGlyPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeuA 173
252 GGAAATTTTTCCTCCCAAAATGCCAACATTCAAAAAGAAATACACTGCT 301
174 IleGlnIlePheSerPheValAlaGluPheSerValProLeuLeuIlePh 190
302 ATACGATTTTCTCTTTTGTGCTGAGTTCTCAGTGCATTCCTTACTCT 351
190 eLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArg 207
352 CCTTTTTCCTTTTGGCTTGTGATTTTCTCTGAGGAGGACACCCGGC 401
207 ImeLysArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaPro 223
402 AAATGAGAAACACAGTGGCCGCGACAGAGGTCTCTGGCAGGGGTGCACC 451
224 IleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyPheSerH1 240
452 ATCAGCGGTGCTGTCTATCTCTCTCTGATCTCTGATCTCTACTCTCC 501
240 sCyMetIleLysValPheLeuSerSerLeuLysPheHisIleLeuArgP 257
502 CTGCATGATTAAGTTTTCCTCTCTCTCAAGTTTTCACATCAGAAGT 551
257 heLePheLeuPhePheIleLeuValIleGlyIleTyProSerGlyHis 273

```

```

seq_name: gb_gss4:A0277039

seq_documentation_block:
  LOCUS      A0277039          310 bp      DNA          GSS          22-NOV-1998
  DEFINITION  CITBI-EI-2516A7.TF CITBI-EI Homo sapiens genomic clone 2516A7, DNA
  sequence.
  ACCESSION  A0277039
  VERSION   A0277039.1  GI:3903235
  KEYWORDS  GSS.
  SOURCE     human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 310)
  Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
  Berry K., Granger D., Suh E., Wible C., Shizuya H., Simon M. and
  Venter J.C.
  Use of a random human BAC End Sequence Database for Sequence-Ready
  Map Building
  Unpublished (1998)
  Other GSS: CITBI-EI-2516A7.TR
  Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
  Seq primer: M13-21
  Class: BAC ends.
  location/Qualifiers
  1. 310
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="2516A7"
  /clone_lib="CITBI-EI"
  /sex="male"
  /cell_type="sperm"
  /note="Vector: pBelBAC11, site_1: EcoRI, site_2: EcoRI;
  Caltech Human BAC Library D"
  BASE COUNT      75 a      77 c      69 g      88 t      1 others
  ORIGIN

alignment_scores:
  Quality: 455.00      Length: 102
  Ratio: 4.643      Gaps: 0
  Percent Similarity: 96.078      Percent Identity: 90.196

alignment_block:
  US-09-510-332-1 x A0277039
  Align seg 1/1 to: A0277039 from: 1 to: 310

124  TGAATTTGCTGCTTCATCCCTTGATTCGATATGATATACCCCTCTCGACAC 601
274  SerLeuLeuLeuLeuLeuGlyAsnProLySLeuLyGlnAsnAlaLyS 269
602  TCTCTCATCTTAATTTTAGCAAAATCCTAATATGAAACAAATGCAAAA 649

seq_name: gb_gss4:A0277039

```

```

103 GGAATTTCCTCCCAAAATGCACAAATTCAAAAGAAAGATACACTGGCT 152
174 IlegInIIepheserPheValAlaGluPheSerValProLeuLeuIlePh 190
153 ATACAGATTACTCTCTGTGTGCTGAGTTCTCAGTGCATTCCTATTCCTT 202
190 eleuPheAlaValLeuLeuLeuIlePheSerLeuClValrghIshPthArg 207
203 CCTTTTCTCTGATATGCTCTAGATTATCTCTGTGGGGGAGACACCCGGC 252
207 lmeFArgAsnthrValAlaGlySerAryValProGlyArgGlyAlaPro 223
225 AAATGAGAAACACACAGCGCGCCGACAGAGTTCTTCGCGAGGGGTGCACCC 302
224 lIeSer 225
303 ATTCAGC 308
seq_name: gb_est7:AA416581
seq_documentation_block:
LOCUS AA416581 463 bp mRNA EST 16-OCT-1997
DEFINITION z005e04.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730974
3', mRNA sequence.
ACCESSION AA416581
VERSION AA416581.1 GI:2077515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 463)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theisling,B., White,Y., Wyllie
,T., Waterston,R. and Wilson,R.
Washu-Merck EST Project 1997
unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -41ml3 fwd. ET from Amer sham
High quality sequence stop: 413.
Location/Qualifiers
1..463
/organism="Homo sapiens"
/db_xref="GDB:3927764"
/db_xref="taxon:9606"
/clone="IMAGE:730974"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
-CTTTCACATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTTCTTTTCTTTT
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldi."
BASE COUNT 140 a 109 c 93 g 121 t
ORIGIN
alignment_scores: 406.00 Length: 83

```

Ratio: 4.892 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 97.590  
Alignment block:  
US-09-510-332-1 x AA416581/rev ..

Align seg 1/1 to reverse of: AA416581 from: 1 to: 463

```

14 ValIleGlnPheLeuLeuGlyIlePheThrAsnGlyIleIleValVala 30
   |||||
265 GTGATACAAATTTCTTCTGGGATTTTCACAAATGATCATGTTGCTGT 216
   |||||
30 1AsnGlyIleAspLeuIleLysSHSArgLysMetAlaProLeuAspLeu 47
   |||||
215 GAATGGCAATTCCTTGATCAGCAGCAAGAAATCGCTCGCATCTCC 166
   |||||
47 euleuSerCysLeuAlaValSerArgIlePheLeuGlnLeuPheIlePhe 63
   |||||
165 TTCTTCTTCTGCTGGCACTTCTAGAAATTTTCTGCACTGTTCACTTC 116
   |||||
64 TyrValaValIleValIlePhePheIleGlnPheIleMetCysSerAl 80
   |||||
115 TACGTTATGTGATGTATCATCTTTCATAGAAATCATCATGTGTCTGC 66
   |||||
80 aAsnCysAlaIleLeuLeuPheIleAsnGlnLeuGlnLeuTyrLeuAla 96
   |||||
65 GAATGTCAATTCCTTATTTATTAATGAATGGAATCTTGCTTGCC 17
   |||||

```

seq\_name: gb\_gsa9:AO616191

seq\_documentation block:

LOCUS AO616191 498 bp DNA GSS 15-JUN-1999  
DEFINITION HS\_5148\_A2\_F05\_SpeR RPrCT-11 Human Male BAC library Homo sapiens  
genomic clone Plate=724 Col=10 Row=K, DNA sequence.

ACCESSION AO616191 GI:5077555  
VERSION AO616191.1  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS 1 (bases 1 to 498)  
Mahalax, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and  
Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
Contact: Mahalax GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

JOURNAL Library availability, please contact Pieter de Jong  
MEDLINE (pieterdejong.med.bufo.fo.edu). Clones may be purchased from  
COMMENT BACPAC Resources (http://bacpac.med.bufo.fo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 724 row: K column: 10  
Seq primer: SP6  
Classes: BAC ends  
High quality sequence stop: 498.

## FEATURES

source

1..498  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate=724 Col=10 Row=K"  
/clone\_lib="RPrCT-11 Human Male BAC Library"  
/sex="male"

/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"  
BASE COUNT 120 a 104 c 112 g 155 t 7 others  
ORIGIN

alignment\_scores:  
Quality: 197.00 Length: 122  
Ratio: 2.402 Gaps: 5  
Percent Similarity: 67.213 Percent Identity: 39.344

alignment block:  
US-09-510-332-1 x AO616191 ..

Align seg 1/1 to: AO616191 from: 1 to: 498

```

34 AspleuIleLysSHSArgLysMetAlaProLeuAspLeuLeuSerCy 50
   |||||
125 GAATGATCAAAACATAGAAAGCTAATGCCAATTCCTTAAATCTG 174
   |||||
50 sleuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValAsn 67
   |||||
175 CATAGCATGTGTCAAAATGCTGTGAGATGTGTGATGTAACAAAGT 224
   |||||
67 alIleValIlePhePhe.....IleGlnPheIleMetCysSer 79
   |||||
225 TTCTCTGCTGTCTTCTTCACCTCTTACGCAAAATTAATCTATGTGCA 274
   |||||
80 AlaAsnCysAlaIleLeuLeuPheIleAsnGlnLeuGlnLeuTyrLeuAl 96
   |||||
275 GCAATGACGTTCTTGTGATGTTTAACTATCACTATGAGCCATGAGGCTG 324
   |||||
96 aThrTrpLeuGlyAlaPheTyrCysAlaLysValAlaSerValArgHis 113
   |||||
325 CACTTGCCTTCTGTATTTACTGCTCAAGAAATTCAGCTCCACCTACGT 374
   |||||
113 roleuPheIleTrpLeuLysMetArgIleSerLysLeuValProTrpMe 129
   |||||
375 CCGTCTTCTTGTGTAATTCAGCA...TCCAAAGGTAATACCTTGCTG 421
   |||||
129 tIleLeuGlySer...leuLeuTyrValSerMetIleCysVal.PheHis 144
   |||||
422 GCTTCTGGAAGCGGCTGCTGCTGAGCAATTCATCTGTGTGTCAC 471
   |||||
145 SerLysTyrAla 148
   |||||
472 GTAGATTACGCC 483
   |||||

```

seq\_name: gb\_est28:AL037695

seq\_documentation block:

LOCUS AL037695 554 bp mRNA EST 29-FEB-2000  
DEFINITION DK2P56400672.r1 564 (synonym: hibr2) Homo sapiens cDNA clone  
DK2P56400672 5', mRNA sequence.

ACCESSION AL037695 GI:5928268  
VERSION AL037695.3  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 554)  
Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J., and Wiemann, S.

JOURNAL EST (Blum, et al.)  
COMMENT On Jul 7, 1999 this sequence version replaced gi:586546.  
Contact: Blum H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert



```

462 ATTTGATTTCTGCTCTTCCACGATGAGTTTGACGAGGGAATCTTGGC 413
23  TAsnAllylelleValValAsnGlylleAspleullelyshsArgL 40
412  CAACGCCCTTCATTGCTGTGTAAGTTGGATGCTGTAAAGGACGAC 363
40  ysmetalProleuAspleulleSerCysleuAlaValSerArgIle 56
362  CCTTGACACACGTGACATCGACACTGCTGTCTACGATCAGCTGGCTT 313
57  pheleuGlnleuPheleuPheleuValAsnValIleValIlePhePheI 73
312  TTCCCTGCAG.....GGCCTTGCTTGGATGCTAT 281
73  eglunPheleuMetCys..... 78
280  TCAG...CTGGCTGCTTCCACGACGATGAAAGCCACTGAGCCACAAC 234
79  ..SerAlaAsnCysAlaIleleuPheleuPheleuAsnGlylleuLeuTrp 94
233  ACCAAGCCATCCTCCTCCTCGATGATTCGAAACCAAGTACGCTCTCG 184
95  leuAlaMetTrpPheGlyValPheTyrCysAlaIysValAlaSerValAr 111
183  CTGGCTGCTGCTCCTCAGTCTCTCTACTGCTCCAAAGATGTCCTCTCTC 134
111  gHis.....ProleuPhele.....TrpleuLysMetArgIleSerL 124
133  TCACACCTTTCACATCCATGACGACGCTGGCTCCAGGAGATTTCTTC 84
124  yslleuValProTrpMetIleleuGlySerleuLeuTyrValSerMetIle 140
83  AGATGCTTCTAGTGTCTTCTCTCTCCTCATCTGACATGACCTGCTTGT 34
141  .....CysValPheHisSer 145
33  TGGGACTTTTTCACATCTCCTCTCC 7

seq_name: gb_gss21:AZ416308

seq_documentation_block:
LOCUS      AZ416308      656 bp      DNA      GSS      03-OCT-2000
DEFINITION 1M0191D04R Mouse 10kb plasmid UUCG1M library Mus musculus genomic
ACCESSION  AZ416308
VERSION    AZ416308.1 GI:10540321
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 656)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tinger,A., von Niederhausen,A.
            and Wright,D., Weiss,R. R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
JOURNAL    Contact: Robert B. Weiss
COMMENT    University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std. Error: 0.00
            Plate: 0191 row: D column: 04
            Seq primer: CACACGAGAAACACGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 656.

```

```

FEATURES
    source
        Location/Qualifiers
            1..656
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUCG1M0191D04"
                /clone_11b="Mouse 10kb plasmid UUCG1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: pMD42uv; Purified genomic DNA from M.
                musculus C57BL/6J (male). Was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptor DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pMD42 (q147321419b|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptor mouse DNA was annealed to
                adaptor vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT      226 a      117 c      119 g      194 t
ORIGIN

alignment_scores:
    Quality:      169.50      Length:      113
    Ratio:        2.230      Gaps:      1
    Percent Similarity: 67.257      Percent Identity: 38.053

alignment_block:
US-09-510-332-1 x AZ416308/rev ..

Align seg 1/1 to reverse of: AZ416308 from: 1 to: 656

188  leuIlePheleuPheAlaValleuLeuIlePheSerleuIlyArgH 204
    :::::::::::::::::::: ::::::::::::::::::::
587  ATGATATGTTAATGTGATGTTCTCTATATCATTTCACTTGAGACA 538
204  sTrpArgGlnMetArgAsnThrValAlaGlySerArgValProGlyArg 221
    I :::::::::::::::::::: IIII ::
537  TCCGAGGACAGATGAAATCAATTAATAGATTACAGATCTCCACAG 488
221  lValAlaProIleSerAlaLeuSerIleleuSerPheleuIleuTyr 237
    :: :::::::::::::::::::: ::::::::::::::::::::
487  AAGTTGATGACACATTAAGATTATGCTTTATATTCCTTTT 438
238  PheSerHisCysMet...IleIysValPheleuSerSerleuPheH 253
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
437  ATATTGCATTTCATGGGTATTCACATTAATGTAATTGCTGTATATCCC 388
253  sIleArgArgPheIlePheleuPheIleuValIleGlyIleTyrP 270
    :::::::::::::::::::: :: ::::::::::::::::::::
387  AGAAGAACACTTGTATTTCATGTTGTTGACACTGCATTCATCTATC 338
270  roSerGlyHisSerleuIleleuIleGlyAsnProIysleuysGln 286
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
337  CCGGCTGCACATCATTATCTTAATTCTAGCAAGAGTGGCTGAAGCG 288
287  AsnAlaIysLysPheleuLeuHisSerIysCysGln 299
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
287  TGCTGTGTAATGATCTGCACTATTAAGTGTCTGTAG 249

seq_name: gb_gss10:AQ711250

seq_documentation_block:

```



```

/clone="RPC1-23-399P19"
/clone_lib="RPC1-23"
/sex="Female"
/lib_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1:
ECORI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      192 a      114 c      101 g      237 t
ORIGIN

alignment_scores:
  Quality: 150.00      Length: 181
  Ratio: 1.456      Gaps: 7
  Percent Similarity: 56.906      Percent Identity: 30.387

alignment_block:
US-09-510-332-1 x AZ078264 ..

Align seg 1/1 to: AZ078264 from: 1 to: 644

129 MetIleuGlySerLeuLeuTyValSerMetIleCysValPheHisSe 145
   ::::::::::::::::::::: ::::: ::::: :::
13 ATTGTCCTAGGCACACATCATGATTTCAATTTTCATATTTCATGCA 62
   ::::::::::::::::::::: ::::: ::::: :::
145 rlystyValagIyPheMetValProTyrPheLeuArgysPhe.....P 160
   : ::::: ::::: ::::: :::::
63 AATGATACCTAAT.....AATTATATACAAAGACCAAAAT 100
   ::::::::::::::::::::: ::::: ::::: :::
160 heSerGlnAsnAlaThrIleGlnIySGIuAsPThrLeuAlaIleGlnIle 176
   ::::::::::::::::::::: ::::: ::::: :::
101 TGGACAAACACACACATTCCTGTTTAGATACCTAAGT..... 141
   ::::::::::::::::::::: ::::: ::::: :::
177 PheSerPheValAlaGluPheSerVal.....ProLeuLeuIlePheLe 191
   ::::::::::::::::::::: ::::: ::::: :::
142 ..GGTTTCCTAGTCTACCATAGCCTACAAATGGGATTCATTTCTT 188
   ::::::::::::::::::::: ::::: ::::: :::
191 uPheAlaVal.....LeuLeuLeuIlePheSerLeuGlyA 203
   ::::: ::::: ::::: :::::
189 TTTTATATAGTCTCGACCTCATTTCTTTTAAATCTTCTCTTTATGGA 238
   ::::::::::::::::::::: ::::: ::::: :::
203 rGHisThArgGlnMetArgAsnThrValAlaGlySerArgValProGly 219
   ::::: ::::: ::::: :::::
239 GGCACCTTAGAGAGATGAACTACAGGCGCATACATACCAAGACATAAGC 288
   ::::::::::::::::::::: ::::: ::::: :::
220 ArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheLeu..... 234
   ::::: ::::: ::::: :::::
289 ACAGAAAGCACATTAAGCTATGAAACSTAGATGATGCTATTCCTTTGTT 338
   ::::::::::::::::::::: ::::: ::::: :::
235 ....IleLeuTyrPheSerHisCysMetIleLeuValPheLeuSerSerL 250
   ::::::::::::::::::::: ::::: ::::: :::
339 CTTCATCATATATATATATACCAACATTAATGCTTATGTGGCAAGCTCCA 388
   ::::::::::::::::::::: ::::: ::::: :::
250 euLySPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIle 266
   ::::: ::::: ::::: :::::
389 TTTCTTGACAAAGTGCTGSCAAATTTTCTTTTAACTA.....ATA 432
   ::::::::::::::::::::: ::::: ::::: :::
267 GlyIleTyrProSerGlyHisSerLeuIleLeuLeuGlyAsnProGly 283
   ::::: ::::: ::::: :::::
433 TTTCTGATTTATCTGTTCATCTTTCTTCTGTGTTTATGGAACACAA 482
   ::::::::::::::::::::: ::::: ::::: :::
283 sLeuIySGIAsnAlaLyAlaLySPheLeuLeuHisSerLeuCys 297
   ::::: ::::: ::::: :::::
483 ATTGAATGACATTCACGATGATGAGAAAGCTGTGTGT 525
   ::::::::::::::::::::: ::::: ::::: :::

seq_name: gb_est66:BE033430

seq_documentation_block: 1176 bp      mRNA      EST      07-JUN-2000
LOCUS      BE033430

```

```

DEFINITION      ME01F08 ME Mesembryanthemum crystallinum cDNA 5' similar to
POLYADICUitin, mRNA sequence.
ACCESSION      BE033430
VERSION        BE033430.1 GI:8328439
KEYWORDS
SOURCE
ORGANISM
Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
Caryophyllales; Alzooceae; Mesembryanthemum.
REFERENCE
1 (bases 1 to 1176)
Bohner, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferreira,
H., Kawasaki, S., McColough, A., Michalowski, C.B., Palacios, C.,
Scara, G., Wheeler, M. and Zepeda, G.R.
Functional Genomics of Plant Stress Tolerance
Unpublished (2000)
Contact: Michalowski, C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: chneu.arizona.edu
An open reading frame exists.
FEATURES
source
1..1176
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_lib="ME"
/tissue_type="roots"
/dev_stage="5-6 weeks"
/note="grown in hydroponics, stress 400 mM NaCl (in 0.5
hoagland's) 6 h stress"
BASE COUNT      129 a      324 c      65 g      658 t
ORIGIN

alignment_scores:
  Quality: 148.00      Length: 307
  Ratio: 0.902      Gaps: 13
  Percent Similarity: 53.420      Percent Identity: 25.081

alignment_block:
US-09-510-332-1 x BE033430 ..

Align seg 1/1 to: BE033430 from: 1 to: 1176

6 LeuIleTyrPheLeuLeuAlaValIleGlnPheLeuGlyIlePhe 22
   ::::: ::::: ::::: :::::
324 CTCTTCATATATTTCTATATGTTCTTTATGCTTTTCTTTCTTTCTTT 373
   ::::::::::::::::::::: ::::: ::::: :::
22 ethAsnGlyIleIleValAlaValAsnGlyIleAsPLeuIleLysAsn 39
   ::::: ::::: ::::: :::::
374 TTCATTC..... 380
   ::::::::::::::::::::: ::::: ::::: :::
39 rGlySMetAlaProLeuAsPLeuLeuLeuSerCysLeuAlaVal.SerAs 55
   ::::::::::::::::::::: ::::: ::::: :::
381 .....CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 416
   ::::::::::::::::::::: ::::: ::::: :::
55 gIlePheLeuGlnLeuPheIlePheTyValAsnValIleValIlePheP 72
   ::::: ::::: ::::: :::::
417 TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 466
   ::::::::::::::::::::: ::::: ::::: :::
72 he.....IleGluPheIleMetCysSerAlaAsnCys 82
   ::::: ::::: ::::: :::::
467 TTTCTATCTATCTATATATCTTTCTATGCTTTCTTTCTTTCTTTCT 516
   ::::::::::::::::::::: ::::: ::::: :::
83 AlaIleLeuLeuPheIleAsnGlyIleGluLeuTyP.....Ile 95
   ::::: ::::: ::::: :::::
517 TTTCTCTCTCTTTCTATTAATCTTTCTTTCTTTCTTTCTTTCTTTCA 566
   ::::::::::::::::::::: ::::: ::::: :::
95 uAlaThrTrpLeuGlyValPheTyCysAlaLyAlaSerValArgH 112
   ::::: ::::: ::::: :::::
567 TTTCTATTTCTTTCTTTCTTTCTTT.....ATCTCTC 598
   ::::::::::::::::::::: ::::: ::::: :::

```



TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LNCM530 row: 1 column: 02  
High quality sequence stop: 54.  
Location/Qualifiers

## FEATURES

## SOURCE

1. 1544  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3840529"  
/clone\_lib="NIH\_MGC\_21"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: placenta; Vector: pCMV7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 982 a 281 c 222 g 59 t  
ORIGIN

alignment\_scores:  
Quality: 141.50 Length: 303  
Ratio: 0.963 Gaps: 18  
Percent Similarity: 48.515 Percent Identity: 26.403

## alignment\_block:

US-09-510-332-1 x BE734359/rev ..

Align seg 1/1 to reverse of: BE734359 from: 1 to: 1544

```

6 leuileiletyrrpheleuAlaValIleGlnpheleuGlyIleph 22
|||||
1148 CTTGTTTGGTGTGTTTCTGCTCTGTTGAGGGTGGCTGCT 1099
|||||
22 ethrasnGlyIleleValIaValaAsnGlyIleAspIleuIleuShsA 39
|
1098 CCTT.....TTGGTCTTCTTGTG..... 1080
|
39 rGlyMetAlaProLeuAspLeuLeuSerCysLeuAlaVal.SerAr 55
|||||
1079 .....CTCTGTGTGGTCTGCTGCTGCTGCTGCTGCT 1050
|
55 gilepheleuGlnpheleuIlephetyrValaAsnValIleValIleph 72
|||||
1049 TGTCTTTGCTGTCTGTTCTGCTGTTGCTGCTGCTGCTGCTT 1000
|||||
72 heileGlnpheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIle 88
|||||
999 TTTTCTGCTTTTCTT.....TTTTCCTCTTCTTGT 968
|
89 AsnGlnleuGlnleuTrpLeuAlaThrTrpLeuGlyValaPheIleCysAl 105
|||||
967 GGTTCCTCTTTTCTGTTGTTCTGCTGCTGCTGCTGCTGCTG 918
|
105 alyValAlaSerValArgHisProLeuPheIleTrpLeuIleuMetArgI 122
|||||
917 TCTGCTTCT.....TTTTCCTGTTCTGCTGCTGCTG 886
|
122 leSerIleuValProTrpMetIleleuGlySerIleuLeuIleuValSer 138
|||||

```

```

885 TTTCGCTGCTGCTCTTTTCTCTTTTGGGCTGTTTCTTTGCTGT 836
|
139 MetIleCysValIleHisSerIleValAlaGlyPheMetValProIle 155
|
835 GTGCTGCTGTG.....TCTCTCTGTTGTTGTTTCTTCT 798
|
155 eleuArgIlePhePheSerGlnAsnAlaThrIleGlnIleGlyAsnPhrI 172
|
797 TAGTCGTTTCT..... 779
|
172 euAlaIleGlnIlePheSerPheValAlaGlnPheSerValProleu 188
|
778 .....TGTCTTTTCTGCTG.....TTCCTGCTGCTTCTTGT 743
|
188 uIlePheleuPheAlaValleuLeuIlePheSerIleuValArgHisT 205
|
742 TTTGTTTGTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
|
205 hrArgIleMetArgAsnThrValAlaGlySerArgValProGlyArgGly 221
|
702 .....TTGCTTCT..... 694
|
222 AlaProIleSerAlaLeu.....LeuSerIleLeuSerPhele 234
|
693 .....CTTCTTTTGTGTTGTTCTGCTGCTGCTGCTGCTGCT 653
|
234 uIleleuTyrrPheSer.....HisCysMetIleIleuValIlePhe 247
|
652 TGTGCTTTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
|
248 SerSerleuIlePheHisIleArgArgPheIlePheleuPheIlele 264
|
602 TTTTGTCTGTTT.....TTTGTCTTCTCTGCTGCTGCTGCTGCT 556
|
264 uValIleGlyIleTyrrProSerGlyHisSerIleuIleleuGlyA 281
|
555 GATTCGTTTCTCTG.....TCTCTTTTGTGTTGTTGTTGTT 518
|
281 snProIleuIleGlnAsnAlaIleIlePheleuIleuHisSerIleu 297
|
517 TTTGTTTCTTCTG.....TTTTCCTGTTGTTTCTGCTGCTG 483
|
298 Cys 298
|
482 TGT 480
|
seq_name: gb_est79:BF037592
seq_documentation_block:
LOCUS BF037592 1502 bp mRNA EST 10-OCT-2000
DEFINITION 601461160F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864615 5',
mRNA sequence.
ACCESSION BF037592
VERSION BF037592.1 GI:10745870
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1502)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCM/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

```



[illegible]

---

**THIS PAGE BLANK (USPTO)**



```

34  apLeuLleUySHsArGLysMeLaIaProLeuAspLeuLeuSerCys 50
101  ACTTATCAAGACAGAAAAATGGCTCGCGTGGATCTCTCTTCTGTT 150
51  LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyValAsnVa 67
151  CTGGCAGTTTCTAGAAATTTTCTCGCAGTTGTTTCATCTTCTACGTTAA 200
67  ILeValIlePhePheIleGluPheIleMeCysSerAlaAsnCysAlaI 84
201  GATTGTATCTCTCTTCATAGAAATTCATCATATGTTCTCGAAATGTCGAA 250
84  LeLeuLeuPheIleAsnGluLeuGluLeuTyPheAlaIaThTrPheGly 100
251  TTCTCTTTTATTAATAATGGAACTTGGCTTGGCCACATGGCTCGGC 300
101  ValPheTyCysAlaValAlaSerValArgHisProLeuPheIleTy 117
301  GTTTTCTATTGTCGCAAGTTGCCAGCGTCCGTCACCCACTCTTCATCTG 350
117  PLeuTyMeLaArgIleSerLysLeuValProTrpMetIleLeuGlySerL 134
351  GTTGAAGATGAGGATATCCAAGCTGGTCCATGGATGATCTGGGGTCTC 400
134  euLeuTyValSerMetIleCysValPheHisSerLysTyValAglyPhe 150
401  TGCTATATGATATATATGATTTGTTGTTCCATACCAATATAGCAGGGTT 450
151  MetValProTyPheLeuArgLysPhePheSerGlnAsnAlaThrIleG 167
451  ATGGTCCCATCTTCTTAGGAAATTTTCTCCAAAAATGCCACAAATTC 500
167  nLySGluAspThrLeuAlaIleGlnIlePheSerPheValAlaGluPhe 184
501  AAAAGAAATACACTGGCTATACAGATTTTCTTTGTTGTTCTGAGTTCT 550
184  erValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSer 200
551  CAGTGCCTATCTTATCTCTCTTTGCTGTTTGCCTTGATTTTCTCT 600
201  LeuGlyArgHisThrArgGlnMetArgAspThrValAlaGlySerArgVa 217
601  CTGGGAGGACACCCGCAATGAGAAACACAGTGGCGGCGACGACGAGGT 650
217  lProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPhe 234
651  TCTGGCAGGGGTGCACCATCACGCGTTCCTATCTCTGTCCTTCC 700
234  euIleLeuTyPheSerHisCysMetIleLysValPheLeuSerSerLeu 250
701  TGATCTCTACTTCTCTCCACTGCATGATAAAAGTTTCTCTCTCTCTA 750
251  LysPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
751  AAGTTTACATCAGACAGGTTTCATCTTCTGTTCTTCATCTTGATGG 800
267  yIleTyPProSerGlyHisSerLeuIleLeuGlyAsnProLysL 284
801  TATATACCTTCTCTGACACTCTCTCATCTTAATTTTAGGAAATCTTAAT 850
284  euLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCysGlyGln 299
851  TGAACAAATGCAAAAAGTTCTCTCTCCACAGTAAAGTCTGTCTG 897

```

seq\_name: gb-prl:AC003015

seq\_documentation\_block:  
LOCUS AC003015 132953 bp DNA PRI 04-FEB-2000  
DEFINITION Human BAC clone GSI-113H23 from Sp15.2, complete sequence.  
ACCESSION AC003015  
VERSION AC003015.1 GI:2547255  
KEYWORDS HTG.

SOURCE  
ORGANISM human.  
REFERENCE  
AUTHORS Homo sapiens  
TITLE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
JOURNAL Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
1 (bases 1 to 132953)  
AUTHORS Wamsley, P., Kramer, J., Elliott, G., and O'Brien, D.  
TITLE The sequence of H. sapiens BAC clone GSI-113H23.  
JOURNAL Unpublished  
2 (bases 1 to 132953)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (20-OCT-1997) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
3 (bases 1 to 132953)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
4 (bases 1 to 132953)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: saplens@wustl.wustl.edu  
----- Summary Statistics  
Center project name: H\_GSI113H23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University School of Medicine, St. Louis MO.

Mapping information for this clone was also provided by Dr. Michael Lovett, Departments of Otorhinolaryngology, Molecular Biology and Oncology, University of Texas Southwestern Medical Center, Dallas TX.

SOURCE INFORMATION:  
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).  
Cell line: lymphoblastoid

Haplotypes: two  
VECTOR: pBel0BAC  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is GSI-330J10, 200 bp overlap. Actual start of this clone is at base position 1 of GSI-113H23; actual end is at 132953 of GSI-113H23.

This clone contains STS HSC022YA5 (NID:91235481) and HS268ZD9 (NID:9454585).

FEATURES  
Source  
This clone contains polymorphisms with GSI-330J10.  
Location/Qualifiers  
1..132953

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone_id="GS1-113H23"
/map="5p15.2"
/complement(20..194)
/rpt_family="L1"
repeat_region
195..358
/rpt_family="ALU"
complement(359..853)
/rpt_family="L1"
repeat_region
2903..2933
/rpt_family="L1"
complement(3640..3667)
/rpt_family="L1"
repeat_region
complement(3764..3788)
/rpt_family="L1"
repeat_region
complement(3820..4112)
/rpt_family="ALU"
complement(4322..4351)
/rpt_family="L1"
repeat_region
complement(4444..4464)
/rpt_family="L1"
repeat_region
complement(5046..5122)
/rpt_family="L1"
repeat_region
complement(5527..6871)
/rpt_family="L1"
repeat_region
complement(6672..6983)
/rpt_family="ALU"
complement(6985..7389)
/rpt_family="L1"
repeat_region
complement(7459..8612)
/rpt_family="L1"
repeat_region
8604..9143
/rpt_family="L1"
repeat_region
9170..9458
/rpt_family="ALU"
repeat_region
9481..9546
/rpt_family="L1"
complement(9591..9823)
/rpt_family="L1"
repeat_region
complement(9854..9961)
/rpt_family="L1"
repeat_region
11262..11306
/rpt_family="L1"
complement(11563..14264)
/rpt_family="L1"
repeat_region
complement(14607..17154)
/rpt_family="L1"
repeat_region
complement(17883..17902)
/rpt_family="L1"
repeat_region
complement(21565..21588)
/rpt_family="L1"
repeat_region
24509..24814
/rpt_family="ALU"
complement(26967..27276)
/rpt_family="L1"
repeat_region
29009..29051
/rpt_family="L1"
complement(29140..29165)
/rpt_family="L1"
repeat_region
complement(29892..30124)
/rpt_family="L1"
repeat_region
complement(30225..30258)
/rpt_family="L1"
repeat_region
31191..31565
/rpt_family="L1"
repeat_region
31649..32016
/rpt_family="L1"
repeat_region
33991..34027
/rpt_family="L1"
repeat_region
35156..35179

```

---

```

repeat_region
complement(36832..36880)
/rpt_family="L1"
repeat_region
39878..39924
/rpt_family="L1"
complement(41297..41751)
/rpt_family="MER"
repeat_region
complement(44442..44621)
/rpt_family="MER"
repeat_region
complement(44871..44963)
/rpt_family="MER"
repeat_region
complement(45153..45205)
/rpt_family="MER"
repeat_region
complement(45255..45761)
/rpt_family="MER"
repeat_region
complement(45830..45882)
/rpt_family="MER"
repeat_region
complement(45907..46229)
/rpt_family="MER"
repeat_region
46846..51590
/rpt_family="L1"
repeat_region
51617..51660
/rpt_family="L1"
repeat_region
52068..52252
/rpt_family="MER"
repeat_region
52278..52459
/rpt_family="MER"
repeat_region
complement(52291..52521)
/rpt_family="MER"
repeat_region
53342..53635
/rpt_family="MER"
repeat_region
54273..54297
/rpt_family="ALU"
repeat_region
58186..58479
/rpt_family="L1"
repeat_region
complement(58482..58529)
/rpt_family="L1"
repeat_region
58482..58529
/rpt_family="L1"
repeat_region
complement(59606..59897)
/rpt_family="ALU"
repeat_region
61733..62140
/rpt_family="ALU"
repeat_region
62686..62745
/rpt_family="L1"
repeat_region
complement(63757..63785)
/rpt_family="L1"
repeat_region
complement(63786..64077)
/rpt_family="ALU"

```

alignment\_scores:

Quality:	1521.00	Length:	299
Ratio:	5.087	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-09-510-332-1 x AC003015 ..

Align seg 1/1 to: AC003015 from: 1 to: 132953

```

1 MetLeuGluSerHisLeuIleIleTyrrhLeuLeuAlaValIleGlnh 17
|||||
54851 ATGCTAGAGTCACCTCATTCATTTCTTCTGACGTATCAATT 54900
17 eldeuGlyIlephethrasnGlyIleIleValAlaAsnGlyIleA 34
|||||
54901 TCTTCTTGAGATTTTCACAAATGCATTCATGTGTGGATGGCATTTG 54950
34 spleuIleTySHsArgIyMetAlaProLeuAspleuLeuSerCys 50
|||||
54951 ACTTGATCAAGCACAGAAAATGGCTCGGATCCCTTCTTCTTGT 55000
51 leuAlaValSerArgIlephethLeuGlnLeuPhelIlephetyrValasVa 67

```



Align seg 1/1 to: AC026787 from: 1 to: 143719

```

1 MetLeuLysSerHisLeuIleIleTyrPheLeuLeuAlaValIleGlnPhe 17
|||||
4517 ATGCTAGAGCTCAACCTCATATCTATCTATCTCTCTCCAGTACATTA 4566
17 eLeuLeuGlyIlePheThrAsnGlyIleIleValIleValIleAsnGlyIle 34
|||||
4567 TCTCTTGGGATTTTCAACAATGCAATCATGTCGTGTGATGCAATG 4616
34 spleuIleLysHisArgLysMetAlaProLeuAspLeuLeuSerCys 50
|||||
4617 ACTGATCAACACACAAAATGCTCCGCGAGATCTCTCTCTCTCTGT 4666
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValAsnVa 67
|||||
4667 CTGGCAGTTTCTAGAAATTTTCTGACAGTGTTCATCTTCTACGTAATGT 4716
67 IleValIlePhePheIleGlnPheIleMetCysSerAlaAsnCysAlaI 84
|||||
4717 GATGTATATCTTCTCATAGAAATCATCATGTCGTGTGCAATGTGCA 4766
84 leLeuLeuPheIleAsnGlnLeuGlnLeuTyrPheAlaIleThrIleGly 100
|||||
4767 TTCTCTTATTAATAATGAATGAACTTGGCTTGGCCACATGCTCGGC 4816
101 ValPheTyrCysAlaLysValAlaIleSerValArgHisProLeuPheIle 117
|||||
4817 GTTTCATATGCGCCAGAGTTCAGCTCCATCCACCATCTTCTCATCTG 4866
117 PheLysMetArgIleSerLysLeuValProTyrMetIleLeuGlySerL 134
|||||
4867 GTTGAAGATGAGATATCCAAAGCTGCTCCATGATGATCTCGGGGCTCT 4916
134 euLeuTyrValSerMetIleCysValPheHisSerLysTyrAlaGlyPhe 150
|||||
4917 TCTCTATATCTATGATTTGTTTCCATAGCAAAATATCAGAGGTTT 4966
151 MetValProTyrPheLeuArgLysPhePheSerGlnAsnAlaIleThrIle 167
|||||
4967 ATGGTCCCATCTCTCTAAGAAATTTTCTCCAAAATGCCACAAATCA 5016
167 nLysGlnAspThrLeuAlaIleGlnIlePheSerPheValAlaGlnPheS 184
|||||
5017 AAAAGAGATACACTGCTATACAGATTTTCTCTTGTGCTGAGATCT 5066
184 euValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSer 200
|||||
5067 CAGTGCATGCTTATCTCTCTTTTGGCTGTTTGTCTGATTTTCTCT 5116
201 LeuGlyArgHisThrArgIleMetArgAsnThrValAlaGlySerArgTyr 217
|||||
5117 CTGGGAGGACACCCGCAAAATGAGAAACACAGTGGCCGCGCAGAGGT 5166
217 lProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheL 234
|||||
5167 TCTGTGAGGAGGTGACCCATCAGCGCGTGTGCTGTATCTCTCTTCC 5216
234 euIleLeuTyrPheSerHisCysMetIleLysValPheLeuSerSerLeu 250
|||||
5217 TGATCCCTACTCTCTCCACTGCATGATTAAGATTTTCTCTCTCTCTA 5266
251 LysPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
|||||
5267 AAGTTTACATCAGAAAGTTCTCTCTCTCTCTCTCTCTCTCTCTCTG 5316
267 YIleTyrProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProLysL 284
|||||
5317 TATATACCTTCTTGACACCTCTCATCTTAATTTTGAAGAACTTAAT 5366
284 euLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCysGlyGln 299
|||||
5367 TGAACAACAAATGCAAAAAGTTCTCTCTCCACAGTAAGTGTCTGTGAG 5413

```

seq\_name: gb\_hcg13:AC034214

```

seq_documentation_block:
LOCUS AC034214 141747 bp DNA 18-JUL-2000
DEFINITION Homo sapiens chromosome 5 clone CTD-2001E22, WORKING DRAFT
ACCESSION AC034214
VERSION AC034214.4 GI:9256725
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 141747)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
PUBLISHED Unpublished
AUTHORS 2 (bases 1 to 141747)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
COMMENT Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7712070.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 628051
Center clone name: CITB-H1_2001E22
-----
Summary Statistics
Consensus quality: 129482 bases at least Q40
Consensus quality: 137633 bases at least Q30
Consensus quality: 139230 bases at least Q20
Estimated insert size: 128750; agarose-fp estimation
Estimated insert size: 141047; sum-of-contigs estimation
Quality coverage: 5.08 in Q20 bases; sum-of-contigs estimation
Quality coverage: 4.64 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
consists of 16 contigs. Gaps between the contigs
are represented as runs of N. The order of the sizes
of the gaps between them are based on estimates that have
provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 21582: contig of 21582 bp in length
* 21583 21582: gap of unknown length
* 21683 21682: gap of unknown length
* 34006 34005: contig of 12323 bp in length
* 34106 34105: gap of unknown length
* 34106 37103: contig of 3000 bp in length
* 37106 37205: gap of unknown length
* 37205 39674: contig of 2469 bp in length
* 39675 39774: gap of unknown length
* 39775 46528: contig of 6754 bp in length
* 46529 46528: gap of unknown length
* 46629 46628: gap of unknown length
* 46629 52637: contig of 6009 bp in length
* 52638 52737: gap of unknown length
* 52738 57097: contig of 4360 bp in length
* 57197 57197: gap of unknown length
* 57198 58513: contig of 1316 bp in length
* 58514 58614: gap of unknown length
* 58614 73885: contig of 15272 bp in length
* 73886 73985: gap of unknown length
* 73986 92578: contig of 18593 bp in length
* 92579 92678: gap of unknown length
* 92679 103695: contig of 11017 bp in length
* 103696 103795: gap of unknown length
* 103796 106232: contig of 2437 bp in length

```

```

* 106233 106332: gap of unknown length
* 106333 121673: contig of 15341 bp in length
* 121674 121773: gap of unknown length
* 121774 124114: contig of 2341 bp in length
* 124115 124215: gap of unknown length
* 124215 131887: contig of 7673 bp in length
* 131888 141747: gap of unknown length
* 131988 141747: contig of 9760 bp in length.
Location/Qualifiers
1. 141747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2001E22"
/clone_lib="Caltech human BAC library D"
BASE COUNT 41436 a 28372 c 28509 g 41923 t 1507 others
ORIGIN

```

```

alignment_scores:
    Quality: 1170.50      Length: 302
    Ratio: 4.417          Gaps: 4
    Percent Similarity: 87.748    Percent Identity: 84.768

```

alignment\_block:  
US-09-510-332-1 x AC034214 ..

Align seg 1/1 to: AC034214 from: 1 to: 141747

```

1 MetLeuGluSerHisLeuIleIleTyrPheLeuLeuAlaValIleGlnPhe 17
|||||
39158 ATGCTAGAGTCTCAGCTATATATCTATTTCTCTGAGGATACATTTT 39207
17 eLeuLeuGlyIlePhePheHisGlnGlyIleIleValIleValIleAsnGlyIle 34
|||||
39208 TCTTCTGGGATTTTACAAATGGCATCATGTGTGGTGATGATGGCAATTG 39257
34 sPLeuIleIleHisArgIleMetAlaProLeuAspLeuLeuSerCys 50
|||||
39258 ACTGATGATCAACACAGAAAATGGCTCCGCGATCTCTCTCTCTCTCTG 39307
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValAsnVal 67
|||||
39308 CTGGCAGTTTCTACAAATTTTCTCAGATTGTCTTCTTCTTACGTTAATGT 39357
67 lIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysAlaI 84
|||||
39358 GATTGTTATCTTCTTCAATGATCATCATGTGTGTGCGAATTGTGCA 39407
84 lLeuLeuPheIleAsnGlnLeuGluLeuTyrPheAlaIleThrPLeuGly 100
|||||
39408 TTCTCTTATTTATTAATGAAATGGAACTTGGCTTGGCAGATGGCTGGC 39457
101 ValPheTyrCysAlaValIleSerValArgHisProLeuPheIleTyr 117
|||||
39458 GTTCTTATTTGTGCCAGATTGCCAGCTCCGTCACCCACTCTTCATCTG 39507
117 PLeuLysMetArgIleSerIleLeuValIleProTyrMetIleLeuGlySer 134
|||||
39508 GTTGAACATAGAGATATCAAGCTGGTCCCATGATGATGATCTT. GGGCTCT 39556
134 euleuTyrValSerMetIleCysValPheHisSerIleVal. TyrAla. GlyP 150
|||||
39557 TGCATATGATGATCATGATTTGTGTTTTCATAGCAAAACATATCAGAGGAT 39606
150 heMetValProTyrPheLeuArgLysPhePheSerGlnAsnAlaThrIle 166
|||||
39607 TTATGGCCCATCTCTCTTAAGGAAATTTTCTCCAAATGCCACAATTT 39656
167 GLeuGluGlnSprHisLeuAlaIleGlnIlePheSerPheValAlaGluPhe 183
|||||
39657 CAAAGACAGATACACTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 39703

```

```

183 eSerValProLeuLeuIlePheLeuPheAlaValIleLeuLeuIlePheS 200
|||||
39704 .....NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 39728
200 eLeuGlyArgHisThrArgGlnMetArgAsnThrValAlaGlySerArg 216
|||||
39729 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 39778
217 ValProGly. ArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerP 233
|||||
39779 GTTCTGGACAGAGGGTGCACCAATACAGCGCTTCTCTC. ATCTGTGCT 39827
233 heLeuIleLeuTyrPheSerHisCysMetIleLysValIlePheLeuSer 249
|||||
39828 TCTGATCTCTTACTTCTCCACATGATATAAGTTTCTCTCTCTCTCT 39877
250 LeuLysPheHisIleArgArgPheIlePheLeuPheIleLeuValIle 266
|||||
39878 CTAAAGTTTCACATCAGAGGTCATCTTCTGTCTTCATCTCTGTGAT 39927
266 eGlyIleTyrProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProL 283
|||||
39928 TGGTATATACCTTCTGGACATCTCTCATCTTAATTGTAGAAATCTTA 39977
283 ySLeuLysGlnAsnAlaLysLysPheLeuLeuHisSerIleCysCysGln 299
|||||
39978 AATTGAACAACAAATGCAAAAAGTCTCTCTCCACAGTAAAGTCTGTAC 40027
seq_name: gb-ro:AF227149

```

```

seq_documentation_block:
LOCUS AF227149 1008 bp DNA ROD 18-MAR-2000
DEFINITION Mus musculus candidate taste receptor t2R19 gene, complete cds.
ACCESSION AF227149
VERSION AF227149.1 GI:7262644
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1008)
Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekar, J., Ryba, N.J. and
Zuker, C.S.
A novel family of mammalian taste receptors
Cell 100 (6), 693-702 (2000)
20222571
2 (bases 1 to 1008)
Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekar, J., Ryba, N.J. P.
and Zuker, C.S.
Direct Submission
Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
Drive, Bethesda, MD 20892, USA

```

```

FEATURES
    source
        1. 1008
            /organism="Mus musculus"
            /strain="129/SvJ"
            /db_xref="taxon:10090"
            1. 1008
                /note="G protein-coupled receptor"
                /codon_start=1
                /product="candidate taste receptor t2R19"
                /protein_id="AA043922.1"
                /db_xref="GI:7262645"
                /translation="MMEGHNLFELVVVQFLGVNGLIVVNAIDLIMKKMAPL
                DLFLCATSRILQLCLFQLGSLVHRTLEADNVTVEYIINELSLWFAWLGVE
                YKATIAIIPHLFMLKMRISRLVPMILASVYVVTPEIHSRETSELPKQIFSR
                SKNTIRVPAAHATLSVFGSLTLPILIPYVAALLLSKMNHSRQRTVNGRREP
                HALVASLSTSELILSLSDMAVVLCTGCLHFGSKTFACLLVIGMTGPSLSIVLI
                LGNFKLNRNKTFTVHCKCHCARAWTSNRPDLVPATHSANKTSKSCSEACIMP
                S"
    CDS
        224 a 252 c 209 g 323 t

```

```

BASE COUNT 224 a 252 c 209 g 323 t
ORIGIN

```

## alignment\_scores:

Quality: 782.00 Length: 299  
 Ratio: 3.218 Gaps: 3  
 Percent Similarity: 81.271 Percent Identity: 51.505

alignment\_block:  
 US-09-510-332-1 x AF227149 ..

Align seg 1/1 to: AF227149 from: 1 to: 1008

```

1 MetLeuGluSerHisLeuIleIleTyRpheLeuLeuAlaValIleGlnP 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 ATGATGGAAGGTCATATGCTCTTCTCTTCTGTCGGTGGTAGTCAGTT 50
17 eLeuLeuGlyIlePheThrAsnGlyIleIleValaValaIasnGlyIleA 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 TTTAACTGGGGCTTGGCAAAATGGCTCTCTGTTGTTGTCATATGCCATCG 100
34 sPLeuIleLysHisArgLysMetAlaProLeuAspLeuLeuSerCys 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 ACTGTATCATGTGGAAGAAATGGCCCCACTGATCTGCTCTTTTTCG 150
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyRValaSnVa 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 CTGGCGACCTTCGGATCATCTCTCAATGTGTATATTTGTCACAGCT 200
67 lIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysAlaI 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 GGGTCATCTCTGTTGGTAGACACACGTTATTTGCTGACATGTTACCT 250
84 lLeuLeuPheIleAsnGluLeuGluLeuTyRLeuAlaThrTyRLeuGly 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 TTGCTACATTAATAACGAAGTCTGCTGTTGCCACATGGCTTGGT 300
101 ValPheTyRValaValaValaValaValaValaValaValaValaVala 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 GTTTTCTACGTGCTCAAGATGCTACCATCCCTCCACCATCTTCTGTG 350
117 PLeuLysMetArgIleSerLysLeuValProTyrMetIleLeuGlySerL 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 GCTGAAGATAGGATATCCAGTGGTGGCCATGGCATCTGTCGATCTG 400
134 eLeuTyRValaSerMetIleCysValaPheHisSerLysTyRAlaGlyPhe 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 TGGTCTATGTAAGCTTACTACTTTCATCCATAGACAGAGACTTCAGAA 450
151 MetValProTyrPheLeuArgLysPhePheSerGlnAsnAlaThr...I 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 CTTCCTTAAGCAAAATCTTATAGCTTTTCTTAAATAATCAACTCGGCT 500
166 eGlnLysGluAspThrLeuAlaIleGlnIlePheSerPheValaIaGluP 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 CAGACCGGCGATGCCACACTACTCTCAGTCTTGTCTT...GGCG 544
183 hEserValProLeuLeuIlePheLeuPheAlaValIleLeuLeuIlePhe 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
545 TCACACATCACATTTCTCATCTTCATCTGCTGTCGCTTGTGTCTCC 594
200 SerLeuGlyArgHisThrArgGlnMetArgAsnThrValaIaGlySerAr 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
595 TCCCTGTGGAAACCAACCGGCGCATAGAG...ACTATGGGGGAGACTAG 641
216 gValProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerP 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
642 GGAACCTTCAGACATGCCCTCTCAGTGGCATGCTCTCCATTCCTGCAT 691
233 hELeuIleLeuTyRPheserHisCysMetIleLysValaPheLeuSerSer 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
692 TCCATCTCTATCTCTCCCATGACATGATGAGCTTCTCATCTGATCACC 741
250 LeuLysPheHisIleArgArgPheIlePheLeuPheIleLeuValI 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq\_name: gb\_ro:AF227140

seq\_documentation\_block:

```

742 CAAGCCTCCACTTTGGAGAGAACCTTGCATTCTGCTTATGTTAT 791
266 eGlyIleTyRProSerGlyHisSerLeuIleLeuGlnLysProL 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
792 TGTATGTATGACCCCTCTTACACTGATTTGCTTAAATTTAGAAACCTTA 841
283 ySLeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCysCys 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
842 AGCTGAACGAATGCAAAACGTTTCATTGTCATTGTCACTGATGTTGT 888

```

LOCUS AF227140 1008 bp mRNA ROD 18-MAR-2000  
 DEFINITION Rattus norvegicus candidate taste receptor T2R1 mRNA, complete cds.  
 ACCESSION AF227140  
 VERSION AF227140.1 GI:7262626  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 1008)  
 Adler,E., Hoon,M.A., Mueller,K.L., Chandrashekar,J., Ryba,N.J. and  
 Zuker,C.S.

TITLE A novel family of mammalian taste receptors  
 JOURNAL Cell 100 (6), 693-702 (2000)  
 MEDLINE 20222571

REFERENCE 2 (bases 1 to 1008)  
 Adler,E., Hoon,M.A., Mueller,K.L., Chandrashekar,J., Ryba,N.J.P.  
 and Zuker,C.S.

TITLE Direct Submission  
 JOURNAL Submitted (21-JUN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center  
 Drive, Bethesda, MD 20892, USA

FEATURES  
 source location/qualifiers  
 1..1008  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"

## CDS

```

1..1008
/note="G protein-coupled receptor"
/codon_start=1
/product="candidate taste receptor T2R1"
/protein_id="AAFA3913.1"
/db_xref="gi:7262627"
/translation="MEGHILFPLVAVVVOFVGLANGLIVVVAIDLIMKKMAPL
DLFLCLATSRITLIQLCLPRQOLPSVLRHTLEEDNITFPIINELISLAPATMGVF
YCAITATIPHLFLMLKMRISRLVPMILSGVLYIITTPHSHRETSALIKPFISLF
PKNATOVGTGHALTSLVLDGLPLFLFVAVLLISLMNTSRMRNTPGRTREYSG
HAHISAMLSILSFILYLSHYMAVVLSTOVLYGSRFVFCILVIGMYSIHISVLI
LGNPKLRKNAKMFIVHCKCHCTRAWVTSRPSRLDLPVPTHPNSANKTSCSEACIMP
S"

```

BASE COUNT 233 a 236 c 205 g 334 t  
 ORIGIN

## alignment\_scores:

Quality: 767.00 Length: 299  
 Ratio: 3.156 Gaps: 4  
 Percent Similarity: 81.271 Percent Identity: 51.839

alignment\_block:  
 US-09-510-332-1 x AF227140 ..

Align seg 1/1 to: AF227140 from: 1 to: 1008

```

1 MetLeuGluSerHisLeuIleIleTyRpheLeuLeuAlaValIleGlnP 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 ATGATGGAAGGCGCATATACCTCTTCTTGTGTTGATGTCAGTGT 50
17 eLeuLeuGlyIlePheThrAsnGlyIleIleValaValaIasnGlyIleA 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 TGTACTGCGGCTTGGCAAAATGGCTCTGTTGTTGTCATGCTATTCG 100

```

```

34 9pLeuileuylshsarglysmetAlaProleuAspLeuLeuLeuSerCys 50
|||||
101 ACTGATGATGTCGGAAGAAATGCCCGCTTGATCTGCTTATATTTTGC 150
51 LeuAlaValSerArgTLeuPheLeuGlnLeuPheIlePheTyrValAsnVa 67
|||||
151 CTGGGCACTTCTCGGATCATCTCGATATATGTTATGTTGCACAAAT 200
67 11LeuAlaIlePhePheIleGluPheIleMetCysSerAlaAsnCysAlaI 84
|||||
201 GGTCTATTCTCTTGGTGAGACACATTATTTAGGACCAATATTACT 250
84 1eLeuLeuPheIleAsnGlnLeuGlnLeuTyrPheLeuAlaThrTyrLeuGly 100
|||||
251 TTGCTCTTCATCAATGACAGTACGACTTGTGTTGCTACACAGCTCCGT 300
101 ValPheTyrCysAlaValAlaSerValAlaArgHisProLeuPheIleTyr 117
|||||
301 GTTTTCTACTGTCGCAAGATTGCTACACTTCTCCACCCACTCTTCTGTG 350
117 PheuLysMetArgIleSerLysLeuValProTyrMetIleLeuGlySerL 134
|||||
351 GGTGAAGATGAGGATATCCAGCTTGCTACATGCTGATCCGGAATCTG 400
134 eulLeuTyrValSerMetIleCysValPheHisSerLys...TyrAlaGly 149
|||||
401 TGCTTATGATGATATTACTACTTTCATCCATACAGACAGACTTCACCA 450
150 PheMetValProTyrPheLeuAlaGlyPheSerHisAsnAlaThrI 166
|||||
451 ATCCCTTAACCAATTTTATA...AGCCTTTTCTCTTAAATATGCACT.. 495
166 eGlnLysGlnAspThrLeuAlaIleGlnIlePheSerPheValAlaGluP 183
|||||
496 CAAGTCGGAACAGGCGCATGCCACACTCTCATGCTGCTGCTTGCGC 544
183 heserValProLeuLeuIlePheLeuPheAlaValLeuLeuIlePhe 199
|||||
545 TCACCTGCTCCGCTGTCATCTTACTGTTGCTGTTCTGCTTGTGATATAC 594
200 SerLeuGlyArgHisThrArgGlnMetArgAsnThrValAlaGlySerAr 216
|||||
595 TCCTGTGGAATTTATAGCAGCAGATGAGC...ACTATGAGACGACACAG 641
216 gValProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerP 233
|||||
642 GGAGATATGCGGACATGCTCATCATGCAATGCTGCTCATTTCTATCAT 691
233 heLeuIleLeuTyrPheSerHisCysMetIleLysValPheLeuSerSer 249
|||||
692 TCCATCATCTCTATCTCTCCACATACATGAGGCTGTTCTGATCTACT 741
250 LeuLysPheHisIleArgArgPheIlePheLeuPhePheIleValI 266
|||||
742 CAAGTCCCTACTCTGGAAGCAGAACTTTGTATTTCTGCTACTAGTTAT 791
266 eGlyTLeuTyrProSerGlyHisSerLeuIleLeuIleLeuGlnAsnProL 283
|||||
792 TGGATGATGACCCCTCAATACACTGATTTGTTAATTTTATGGAATCCCTA 841
283 yLeuLysGlnAsnAlaValLysLysPheLeuLeuHisSerLysCysCys 298
|||||
842 AGCTGAACGAAATGCAAAATGTTTCAATTCATTTGTAAGTCTTGT 888

```

KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 180438)  
 Waterston,R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 180438)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (12-NOV-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Jul 15, 2000 this sequence version replaced g1:8954416.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Project Information  
 Center project name: H\_NH0472F21  
 Summary Statistics  
 Sequencing vector: M13; 78%  
 Sequencing vector: plasmid; 22%  
 Chemistry: Dye-terminator Big Dye; 22% of reads  
 Chemistry: Dye-terminator Big Dye; 22% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 173153 bases at least Q40  
 Consensus quality: 175380 bases at least Q30  
 Consensus quality: 176853 bases at least Q20  
 Insert size: 193000; agarose-fp  
 Insert size: 178738; sum-of-contigs  
 Quality coverage: 4.55 in Q20 bases; agarose-fp  
 Quality coverage: 5.07 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 18 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

Contig	Length	Quality	Insert Size	Sum-of-contigs
1	1282	contig of 1282 bp in length		
1283	1382	gap of unknown length		
1383	2619	contig of 1237 bp in length		
2620	2719	gap of unknown length		
2720	4393	contig of 1674 bp in length		
4394	4493	gap of unknown length		
4494	10415	contig of 5922 bp in length		
10416	10515	gap of unknown length		
10516	17141	contig of 6626 bp in length		
17142	17241	gap of unknown length		
17242	24045	contig of 6804 bp in length		
24046	24145	gap of unknown length		
24146	32644	contig of 8499 bp in length		
32645	32744	gap of unknown length		
32745	41130	contig of 8366 bp in length		
41131	41230	gap of unknown length		
41231	49280	contig of 8050 bp in length		
49281	49380	gap of unknown length		
49381	60265	contig of 10885 bp in length		
60266	60365	gap of unknown length		
60366	72486	contig of 12121 bp in length		
72487	72586	gap of unknown length		
72587	84443	contig of 11857 bp in length		
84444	84544	gap of unknown length		
84544	93472	contig of 10929 bp in length		
93473	95572	gap of unknown length		
95573	108950	contig of 13378 bp in length		
108951	109050	gap of unknown length		



```

ACCESSION      AC005541
VERSION        AC005541.1
KEYWORDS       GI:4156136
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE          1 (bases 1 to 110626)
JOURNAL        Waterston, R.H.
REFERENCE      The sequence of Homo sapiens clone
AUTHORS        unpublished
TITLE          2 (bases 1 to 110626)
JOURNAL        Waterston, R.H.
REFERENCE      Direct Submission
AUTHORS        Submitted (25-AUG-1998) Genome Sequencing Center, Washington
TITLE          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL        MO 63108, USA
REFERENCE      3 (bases 1 to 110626)
AUTHORS        Waterston, R.H.
TITLE          Direct Submission
JOURNAL        Submitted (14-JAN-1999) Genome Sequencing Center, Washington
REFERENCE      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL        MO 63108, USA
COMMENT        On Jan 14, 1999 this sequence version replaced gi:3907453.
FEATURES       Location/Qualifiers
SOURCE         1..110626
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="RC044L22"
BASE COUNT     35376 a 19647 c 20125 g 35478 t
ORIGIN
alignment_scores:
      Quality: 434.50      Length: 301
      Ratio: 2.099         Gaps: 7
      Percent Similarity: 68.771 Percent Identity: 33.223
alignment_block:
US-09-510-332-1 x AC005541/rev ..
Align seg 1/1 to reverse of: AC005541 from: 1 to: 110626
7   llelelyrphleuleuAlaVallelelnpheuleuglylelpheth 23
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4392 ATTCTTCATATTATCATGATGTCACAGAAATTCCTCACAGGATCACAGT 4343
23 rAsnGlyllelleValValAsnGlylleAspLeuilelshsArgL 40
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4342 AAATGAGATTCTTATCATGTTTAACTGTAATGATGATGATCAAAACATAGAA 4293
40 ysmcAlaProleuAspLeuleuSerCysLeuAlaValSerArgIle 56
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4292 AGCTAAATGCCAAATTCCTTAATGTCCATAGGAGATGCTAAGATT 4243
57 pheleuglnleuphelelphetyrValAsnVallellelphethe.. 72
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4242 GGTGTGGAGATGCGTGAATGTCACAAAGTTTTCCTGCTGTTCTTCC 4193
73 .....llegluphelelmetCysSerAlaAsnCysAlaIleleuL 86
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4192 ACTGCTTATTCACAAATTAATTGTTGTCGACAAATGATGTCCTTTCGA 4143
86 eupheileanglnleuglnleulrleuAlarhtrrpleuglyValrhe 102
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4142 TGTTTTAACTATACGCTATGAGCTTGGTCCACTTGCCTTCTGTATTT 4093
103 tyrcAlaIaYsValAlaSerValArgnIleProleuphelelletrpleu 119
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
4092 TACGCGCTCAAGATTTACGCTTCACATCCTGCTTTTGTGGTGGA 4043
119 smetArgIleSerlyLeuValProtrpmetIleleuglySerleuLeu. 135
   |||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

4042 ATTCAGATCCCAAGTTAATACCTTGCTGCTCTGCGAAGCCTTCTG 3993
136 ..TyValSermetIleCysValrPheHisserlyS..TyValaglyPhe 151
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3992 CCTGTGTAGACATTGCATCTGTGTGTGTCAGGTAGATTACGCTAAATATG 3943
151 tValProTyrrPheleuArglySphPheSerGlnAsnAlaThrIleGlnL 168
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3942 GCAACAGCATGCCCTCAGA.....AACACCACTAAATAA 3908
168 ySGLuAspThrleuAlaIleGlnIlePheSerPheVal.....Ala 181
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3907 AGAGTAAACAAAGATTAAGAAATTTGTCGAAGCTTCTTGTCACTTG 3858
182 GlupheSerValProleuIlelPheleupheAlaValleuLeuIleuL 198
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3857 GCATTAAATATTCTCTAGCCATATTGTGTGATGTCACCTTCTATGTTACT 3808
198 ePheSerleuglyArgnIsthrArglnMetArgAsnThrValAlaGlyS 215
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3807 CATCTCTTTTACAAGCACACTCATCGATGCAACATGATGATTCATGCT 3758
215 eArGValProGlyArgGlyAlaProIleSerAlaLeuSerIleleu 231
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3757 TTAAATATGCCAACACAGAACCCATTAATGATTAATAAACAGTATATA 3708
232 SerPheleuIleleuTyrrPheSerHisCysmetIlelyValrPheleuS 248
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3707 ACATTCCTTTGCTCTTTAATTCTTAT.....TTGCTGCTCTCATGAC 3664
248 rSerleuLySPhelHisIleAlrArgPheIlePheleuphePheIleu. 264
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3663 AAATATACATTTAGTTTACCTTACACAGAACTCACAGCTTCTTATGCTGA 3614
265 ....ValIleGlylleTyrrProSerCylHisSerleuIleleuL 279
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3613 AGCACAATATAGCGACATATCCCTGCGCACCTGCTTAATATATCTTG 3564
280 GLyAsnProLySleuLySGLnAsnAlaLySlySphleuLeuHisserly 296
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
3563 AGTAAATTTAGTTCCACACATCATTTAGAAAGAAATTCCTGCTCAAAA 3514
296 s 296
3513 G 3513
seq_name: gb_ptr4:AF227130
seq_documentation_block:
LOCUS      AF227130          951 bp      DNA          PRI          18-MAR-2000
DEFINITION Homo sapiens candidate taste receptor T2R3 gene, complete cds.
ACCESSION  AF227130
VERSION    AF227130.1  GI:7262606
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      A novel family of mammalian taste receptors
JOURNAL    Cell 100 (6), 693-702 (2000)
MEDLINE    20222571
REFERENCE  2 (bases 1 to 951)
AUTHORS    Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekhar, J., Ryba, N.J. and
            Zuker, C.S.
            A novel family of mammalian taste receptors
            Direct Submission
            Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
            Drive, Bethesda, MD 20892, USA
            Location/Qualifiers
            1..951
            /organism="Homo sapiens"
FEATURES
SOURCE

```



Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
 ----- Summary Statistics -----  
 Center project name: H\_DJ1154E09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send [mailto:sgreen@nhgri.nih.gov](mailto:mailto:sgreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.  
 The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.  
 VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP5-894A10, 200 bp overlap: the clone sequenced to the right is unknown. Actual start of this clone is at base position 103168 of RP5-894A10; actual end is at 82419 of RP5-1154E9.

#### FEATURES

source  
 location/Qualifiers  
 1..82419  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q31.3-q32"  
 /clone="RP5-1154E9"  
 /clone\_1lb="RPCI-5"  
 1..62  
 /rpt\_family="Alu"  
 259..800  
 /rpt\_family="MER1\_type"  
 801..1104  
 /rpt\_family="Alu"  
 1105..1132  
 /rpt\_family="MER1\_type"  
 1134..1434  
 /rpt\_family="Alu"  
 1437..1740  
 /rpt\_family="L1"  
 1830..1865  
 /rpt\_family="MIR"  
 2441..2765  
 /rpt\_family="MER2\_type"  
 2918..3211  
 /rpt\_family="Alu"  
 3226..3272  
 /rpt\_family="AT-rich"  
 misc\_feature  
 3239..3594

misc\_feature  
 3365..3878  
 /note="match to EST C15294 (NID:91570001)"  
 misc\_feature  
 3498..3878  
 /note="match to EST AA649071 (NID:92575500) ns42h02.s1"  
 STS  
 3510..3756  
 /db\_xref="GI:435384"  
 misc\_feature  
 3587..3874  
 /note="similar to EST D57847 (NID:9664469)"  
 repeat\_region  
 4204..4293  
 /rpt\_family="L2"  
 4543..5520  
 /rpt\_family="Retroviral"  
 5535..8698  
 /rpt\_family="Retroviral"  
 8789..9520  
 /rpt\_family="Retroviral"  
 9522..9556  
 /rpt\_family="Retroviral"  
 11462..11528  
 /rpt\_family="AT-rich"  
 11735..12045  
 /rpt\_family="Alu"  
 12587..12789  
 /rpt\_family="MIR"  
 12884..13241  
 /rpt\_family="MaLR"  
 13244..13907  
 /rpt\_family="MaLR"  
 13902..14230  
 /rpt\_family="MaLR"  
 14247..14415  
 /rpt\_family="L2"  
 14427..15098  
 /rpt\_family="L2"  
 /rpt\_family="L2"  
 15151..15195  
 /rpt\_family="AT-rich"  
 15213..15353  
 /rpt\_family="Alu"  
 15546..16058  
 /rpt\_family="MER73-group"  
 16279..16492  
 /rpt\_family="MaLR"  
 18279..18327  
 /rpt\_family="purine-rich"  
 18911..19019  
 /rpt\_family="MIR"  
 19567..19706  
 /rpt\_family="MIR"  
 19707..19753  
 /rpt\_family="(CA)n"  
 19754..19835  
 /rpt\_family="MIR"  
 20529..20588  
 /rpt\_family="L2"  
 20778..21195  
 /rpt\_family="Retroviral"  
 21265..21293  
 /rpt\_family="(CA)n"  
 21294..21896  
 /rpt\_family="L2"  
 21969..22393  
 /rpt\_family="L2"  
 22470..22492  
 /rpt\_family="MIR"  
 22493..22865  
 /rpt\_family="MaLR"  
 22866..22954  
 /rpt\_family="MIR"  
 23042..23082  
 /rpt\_family="AT-rich"  
 23385..29503  
 /rpt\_family="L1"



```

/translaction="MESPADNIFILITSEFIILGNGYIALVNIWMJMKKKISIV
DYILNLNVIARICLISVMVNGIVIVANPVTYKNNKOIYIFFWTFANLNMITTC
LNVPEYELKISSSSHPLELWKWIDMYVHMLIGCEPILSLVLAIVLSCDYREHA
IAKKRNITTEHVSKEIPTEPLENLFAIVPFIYSLISFELSLMRHTQIKLY
ATGSRDPSSTEVHRAIKTMTSEFFELVYISSILMTFYSYLMTRYKLAVFEGIAAIL
VPLGHSLILVILNNKLRQTFEVRMLTCKRIACMI"
<1..>930
/product="candidate taste receptor T2R8"
primer_bind 907..930
BASE COUNT 279 a 171 c 161 g 319 t
ORIGIN

alignment_scores:
  Quality: 376.00      Length: 320
  Ratio: 2.000         Gaps: 10
  Percent Similarity: 58.750   Percent Identity: 31.562

alignment_block:
US-09-510-332-1 x AF227134 ..

Align seg 1/1 to: AF227134 from: 1 to: 930

      8  ILETyrPheLeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThrAs 24
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      22  ATCTTTAATAATCTTAATACTGAGCAATTCATACGAAATATATGGGAA 71
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      24  nGlyIleIleValIaValAsnGlyIleAspLeuIleIleYshAsArgLysM 41
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      72  TGGATACATTCGACACTGACACTGATTCGATTCGATTAAGAAAGAAAGA 121
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      41  eAlaIleProLeuAspLeuLeuLeuSerCysLeuAlaValSerArgIlePhe 57
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      122  TTTCCACAGCTTGACACTACCTTACCAATTTAGTTATGCCAGAAATTTGT 171
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      58  LeuGlnIlePheIlePheTyrValAsnValIleValIlePhePheIleGl 74
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      172  TTG..ATCAGTGTAAATGCTGTAAATGTCATTTAATAGTACGAACCC 218
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      74  uPheIleMetCysSerAlaAsnCysAlaIleLeuLeu..... 86
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      219  AGATGCTTATACAAATAAATAAACAAGATGTCATTTTACCTTCGGA 268
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      87  ..PheIleAsnGlnLeuGlnLeuTyrPheLeuAlaThrTyrPheGlyValPhe 102
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      269  CATTTGCCCACTACTTAATATGTGTGATTTACCACTGCTGCTTAATGCTTTC 318
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      103  TyrCysAlaIleValAlaSerValArgHisProLeuPheIleTyrPheLeu 119
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      319  TATTTTCTGAAGATAGCCAGTCTCTCATATCCACTTTTCTCTGGCGAA 368
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      119  sMetArgIleSerLysLeuValProTyrPheIleLeuGlySerLeu.... 134
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      369  GTGGAATAATGATATGTGTGTCACGATCGATCGCTGCGATGCTTTCGA 418
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      135  ....LeuTyrValSerMetIleCysVal..... 142
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      419  TTTCTTGTGTGCTCAGCTTATAGCAGCAATAGTACGAGTTGTGATTAT 468
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      143  ...PheHisSer.....LysTyrAla 148
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      469  AGCTTTCATGCAATTGGCAAAACATAAAGAAACATTCTGAATGTTTCCA 518
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      148  aGlyPheMetValProTyrPheLeuArgLysPhePheSerGlnAsnAlaT 165
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      519  TGTGACTAAATACATATCTTT..... 540
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      165  hrlIleGlnLysGluAspThrLeuAlaIleGlnIlePheSerPheValAla 181
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      541  ....GAACTCTTACTCTCTTAACTGTTTGAAT..... 573
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      182  GluPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuLeu 198
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

---

```

574  ....GTCCATTATTGTGTCTACGTATATCAATTTTCTTTAGT 614
198  ePheSerLeuGlyArgHisThrArgGlnMetArgAsnThrValAlaGlys 215
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
615  AAGATCTTATGAGACATACCAAGCAATTAACCTATGCTACCGCA 664
215  eArgValProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeu 231
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
665  GTAGAGACCCCGACAGCAAGTTCATGAGAGCCATTAACATATGACT 714
232  SerPheLeuIleLeuTyrPheSerHisCysMetIleLysValPheLeu 248
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
715  TCATTATCTCTTTTCTTCTTAC.....TATATTTC 749
248  rSerLeuLysPheHisIleArgArgPheIlePheLeu..... 260
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
750  TTCATAT.....TTGATGACCTTTAGCTATCTATGCAAAATACA 790
261  ....PhePheIleLeuValIleGlyIleTyrProSerGly 272
791  AGTTAGCTGTGAGTTTGAGAGATTGCACCAATTCCTACCCCTGGGT 840
273  HisSerLeuIleLeuIleLeuGlyAsnProLysLeuLysGlnAsnAla 289
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
841  CACTCATTTATTTTATTTATTTAATTAATTAATTAACAGCAGACATTTGT 890
289  sLysPheLeu 292
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891  CAGCAATGCTG 900

seq_name: gb_p4:AF227133

seq_documentation_block:
LOCUS      AF227133          957 bp      DNA          PRI      18-MAR-2000
DEFINITION Homo sapiens candidate taste receptor T2R7 gene, complete cds.
ACCESSION  AF227133
VERSION    AF227133.1 GI:7262612
KEYWORDS
SOURCE
  ORGANISM  human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 957)
            Adler,E., Hoon,M.A., Mueller,K.L., Chandrasekhar,J., Ryba,N.J. and
            Zuker,C.S.
            A novel family of mammalian taste receptors
            Cell 100 (6), 693-702 (2000)
TITLE      JOURNAL
MEDLINE    20222571
REFERENCE  2 (bases 1 to 957)
            Adler,E., Hoon,M.A., Mueller,K.L., Chandrasekhar,J., Ryba,N.J. P.
            and Zuker,C.S.
            Direct submission
            Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
            Drive, Bethesda, MD 20892, USA
FEATURES
  source
    1..957
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="12"
    /map="12p13"
    1..20
    1..957
    /note="G protein-coupled receptor; PCR-derived sequence;
    similar to an interval in BAC AC006518"
    /codon_start=1
    /product="candidate taste receptor T2R7"
    /protein_id="AA043906.1"
    /db_xref="GI:7262613"
    /translation="MADKVOITLFLAVGESVGIIGNAFISGLVNCMDWKKRKIASI
    DLITSLAISRICLIVITLDFLIVYPPVYATGKEMRIIDFPFTLTHLSITPAC
    LSTYEPKIGNFPHPLEFIAMKMRIDRIVISILIGCYVLSFISLPATENINADRFRCY
    KAKRKNITNSCRVNTQHSATKFLNLATLFLFCVCLMSFLLITLSLRHNRMQLS
    ATGCRDPSSTBAHVRAKAVISFLLFLIAYLSFLIATSSYFMETELAVIFESIALI

```

```

141 ysvalPheHisSerLysIAspAIA.....Gly 149
      ||||| |||
422 CTGTCTATTATTAAGCCTTCCAGCCACTGACATTTGAAGCGTCATTTCAG 471
      ||||| |||||
150 PheMetValProTyrPheLeuArgLysPheSerGlnAsnAlaThrIl 166
      ||||| :::: :::: :::: ::::
472 TTTTCTGTGAAGGCACAAGAGCAAAACAATTACTTGAGTGGCAGCAT 521
      ||||| :::: :::: :::: ::::
166 eGIInLySGlUASPTInLeuAlleGlInILePheSerPheValAlaGluP 183
      ||||| :::: :::: :::: ::::
522 AATATAACTCAACATGCTCTTACCAGAATTATTTCACCTGGCAAGCG 571
      ||||| :::: :::: :::: ::::
183 HisSerValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePhe 199
      ||||| :::: :::: :::: ::::
572 TG...CTCCCCTTTGTGTGGCTATGATGCCCTTTTCCCTTGATTCCTC 618
      ||||| :::: :::: :::: ::::
200 SerLeuGIyArgHisThrArgLIneTarGSnthrValAlaGlySerAr 216
      ||||| ||||| ||||| :::: :::: :::: ::::
619 FCCCTCGGAGACATATCCAGGCGCAATGCAGCTCAGTGCCACAGGGTCAG 668
      ||||| ||||| ||||| :::: :::: :::: ::::
216 yvalProGIyArgGIyAlAProIleSerAlaLeuLeuSerIleLeuSerP 233
      ||||| ||||| ||||| :::: :::: :::: ::::
669 AGACCCGACGACNAGAACCCCATGTGAGAGCCCTGGAAGCGTGTATTTCCT 718

```

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project information  
Center project name: HMOY  
Center clone name: RP13-81N3  
-----  
Summary Statistics  
Sequencing vector: M13: L08821  
Chemistry: Dye-Primer Bodipy: 12% of reads  
Chemistry: Dye-terminator Big Dye: 86% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 129880 bases at least Q40

NOTE: Estimated insert size may differ from sequence length (see [http://www.hpsc.bcm.tmc.edu/docs/genbank\\_drift\\_data.html](http://www.hpsc.bcm.tmc.edu/docs/genbank_drift_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

**FEATURES**  
**SOURCE**

alignment\_scores:

alignment\_block:  
US-09-510-332-1 x AC016145/rev

57682 AGCTTAATCTACCCCTCAAGTCATTTCATTTATCCTAATACTGCGGACACA 57633

[illegible]

57632 ATAAATTAGACATGCATCTAAAGTGATT 57601

seq\_name: gb\_pr2:AC006518

seq\_documentation\_block: AC006518 173735 bp DNM PRI 01-MAY-1999

LOCUS Homo sapiens 12p13 BAC RPi11-144023 (Roswell Park Cancer Institute

ACCESSION AC006518 Human BAC Library) complete sequence.

VERSION AC006518.17 GI:4713939

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE 1 (bases 1 to 173735)

JOURNAL Muzny, D., Aronson, A.D., Bouck, J., Bunac, C., Chen, Z., Ding, Y.,

AUTHORS Dugan, S., Duhin, J., Forcum, J., Garcia, C., Correll, J. H.,

JOURNAL Kordejski, N., Lau, S., Leal, B., Lee, E., Lichtarge, O., Liu, W.,

AUTHORS Logan, O., Lu, J., Marondel, I., Martinez, C., Merscher, S., Miller, A.,

JOURNAL Montgomery, K., Oswal, G., Pampell, L. R., Parish, B. J., Perez, L.,

AUTHORS Rashid, N. D., Rives, C., Scherer, S. E., Shen, H., Shim, C., Simon, M.,

JOURNAL Vo, Q., Williamson, A., Worley, K. C., Xiang, A. M., Yang, R., Yu, W.,

AUTHORS Zhou, X., Kuchelapatti, R., Nelson, D. and Gibbs, R. A.

JOURNAL Direct Submission

AUTHORS 2 (bases 1 to 173735)

JOURNAL Worley, K. C.

AUTHORS Direct Submission

JOURNAL Submitted (05-FEB-1999) Molecular and Human Genetics, Baylor

AUTHORS College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL 3 (bases 1 to 173735)

AUTHORS Worley, K. C.

JOURNAL Direct Submission

AUTHORS Submitted (29-APR-1999) Human Genome Sequencing Center, Department

JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One

AUTHORS Baylor Plaza, Houston, TX 77030, USA

JOURNAL On Apr 29, 1999 this sequence version replaced gi:4589934.

JOURNAL INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email

annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----

Contig length: 173735

Phrap values in estimate: 167393

Average error rate (BCR-Phrap estimate): 0.000263256

Fraction of Phrap values less than 40 : 0.0395058

Number of consensus changing edits: 3

Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original-Context	Edited-Context
31176	gtgggtgag(n)agagggag	gtgggtgag(g)gagggag
31177	tgggtgag(n)agagggag	tgggtgag(g)gagggag
104155	gagttcact(n)gtgtcttc	gagttcact(c)gtgtcttc

----- Distribution of Quality &lt; 40 Bases -----

#	5	10	15	20	25	30	35	40
10001	*	*	*	*	*	*	*	*
9001	*	*	*	*	*	*	*	*
8001	*	*	*	*	*	*	*	*
7001	*	*	*	*	*	*	*	*
6001	*	*	*	*	*	*	*	*
5001	*	*	*	*	*	*	*	*
4001	*	*	*	*	*	*	*	*
3001	*	*	*	*	*	*	*	*
2001	*	*	*	*	*	*	*	*
1001	*	*	*	*	*	*	*	*
01	*	*	*	*	*	*	*	*

Phrap Value Range

Version: 1.01 gxf.

FEATURES Location/Qualifiers

source

1..173735

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="RP11-144023"

/chromosome="12p13"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

/rpt\_family="MER53"

```

repeat_region      /rpt_family="AT_rich"
                    14232..14649
                    /rpt_family="LIM4"
repeat_region      /rpt_family="LIM4"
                    14652..14742
                    /rpt_family="(TA)n"
repeat_region      15287..15421
                    /rpt_family="L2"
repeat_region      16053..16237
                    /rpt_family="LIMB5"
repeat_region      16238..16272
                    /rpt_family="U2"
repeat_region      16273..17683
                    /rpt_family="LIMB5"
repeat_region      18072..18210
                    /rpt_family="LIM4A"
repeat_region      18679..18844
                    /rpt_family="LIM4"
repeat_region      18895..19193
                    /rpt_family="LIPAL6"
repeat_region      complement(19194..19232)
                    /rpt_family="AT_rich"
repeat_region      complement(19658..20023)
                    /rpt_family="LIPB2"
repeat_region      complement(20437..20931)
                    /rpt_family="LIM4A"
repeat_region      complement(25697..25720)
                    /rpt_family="AT_rich"
repeat_region      26549..26960
                    /rpt_family="LIM4"
repeat_region      27024..27268
                    /rpt_family="LIM4"
repeat_region      27269..27511
                    /rpt_family="MER46"
repeat_region      27793..28087
                    /rpt_family="AluJo"
repeat_region      complement(28088..28123)
                    /rpt_family="(GA)n"
repeat_region      complement(28124..28214)
                    /rpt_family="(TA)n"
repeat_region      28277..28381
                    /rpt_family="LIM4"
repeat_region      28368..28912
                    /rpt_family="LIM4"
repeat_region      28925..29052
                    /rpt_family="MLT1A1"
repeat_region      29201..29267
                    /rpt_family="Alus"
repeat_region      complement(29508..29559)
                    /rpt_family="(TAAA)n"
repeat_region      complement(29560..29858)
                    /rpt_family="AluY"
repeat_region      complement(30369..30396)
                    /rpt_family="AT_rich"
repeat_region      31146..31316
                    /rpt_family="LIP2"
repeat_region      31857..32070
                    /rpt_family="AluJo"
repeat_region      complement(32384..33480)
                    /rpt_family="LIM45"

```

```

alignment_scores:
  quality: 376.00      length: 320
  ratio: 2.000         gaps: 10
  percent similarity: 58.750   percent identity: 31.562

```

alignment\_block:  
us-09-510-332-1 x AC006518/rev ..

Align seg 1/1 to reverse of: AC006518 from: 1 to: 173735

```

8 11eTyPheLeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThrAs 24
|||||:|||||:|:|:|:|||||:|||||:|:|:|:|
5870 ATCTTATATATCTTAATACGTGAGAAATTCATCTAGCAATATATGGGAA 5821

```

```

24 nGlyIleIleValAlaValAsnGlyIleAspLeuIleLysArgLysM 41
   |||:|||||:|||||:|||||:|||||:|||||:|
5820 TGGATACATCTGCACTAGTCAACAGTGTGAGTGAATTAAGAAAAAGA 5771
   |||:|||||:|||||:|||||:|||||:|||||:|
41 eTAlaPProLeuAspLeuLeuSerCysLeuAlaValSerArgIlePhe 57
   |||:|||||:|||||:|||||:|||||:|||||:|
5770 TTTCACAGTTGACATACATCTTACCAATTTAGTTATCCGCAAAATTTGT 5721
   |||:|||||:|||||:|||||:|||||:|||||:|
58 LeuGlnLeuPheIlePheTyGValAsnValIleValIlePheIleG 74
   |||:|||||:|||||:|||||:|||||:|||||:|
5720 TTG...ATCAGTGTAAATGCTTAAATGCAATGTCTAATAGTACTGAACC 5674
   |||:|||||:|||||:|||||:|||||:|||||:|
74 uPheIleMetCysSerAlaAsnCysAlaIleLeuLeu.....86
   |||:|||||:|||||:|||||:|||||:|||||:|
5673 AGATGGTTATATACAAAATAAACACAGATATTCATTTTACCTCTGGA 5624
   |||:|||||:|||||:|||||:|||||:|||||:|
87 ..PheIleAsnGlyLeuGlyLeuTyrPheLeuAlaThrTrpLeuGlyValPhe 102
   |||:|||||:|||||:|||||:|||||:|||||:|
5623 CATTTGCCAAGCTACTTAATATGTGGATTACACCTGCCTTAATGTCTTC 5574
   |||:|||||:|||||:|||||:|||||:|||||:|
103 TyrCysAlaLeuValAlaSerValArgHisProLeuPheIleTrpLeuTy 119
   |||:|||||:|||||:|||||:|||||:|||||:|
5573 TATTTCTGAAAGATAGCAGATTCCTCTCATCCACTTTTCTCTGCTGTA 5524
   |||:|||||:|||||:|||||:|||||:|||||:|
119 sMetArgIleSerLysLeuValProTrpMetIleLeuGlySerLeu...134
   |||:|||||:|||||:|||||:|||||:|||||:|
5523 GTGGAAATATGATATGTGTGCATCTGCCTGCGGATGCTTTGGCA 5474
   |||:|||||:|||||:|||||:|||||:|||||:|
135 ....LeuTyGValSerMetIleCysVal.....142
   |||:|||||:|||||:|||||:|||||:|||||:|
5473 TTTCCTGTGTGGTACGCTTATAGCAGCAATAGTACTGAGTTGTGATAT 5424
   |||:|||||:|||||:|||||:|||||:|||||:|
143 ...PheHisSer.....LysTyGAl 148
   |||:|||||:|||||:|||||:|||||:|||||:|
5423 AGCTTTGACAGCAATGGCCAAACATAAAGAACATTACTGAAATTTGCCA 5374
   |||:|||||:|||||:|||||:|||||:|||||:|
148 aGlyPheMetValProTyPheLeuArgLysPheSerGlnAsnAlaIat 165
   |||:|||||:|||||:|||||:|||||:|||||:|
5373 TGTGAGTAAATATCCACTTT.....5352
   |||:|||||:|||||:|||||:|||||:|||||:|
165 hTrIleGlnLysGlnAspThrLeuAlaIleGlnIlePheSerPheValAla 181
   |||:|||||:|||||:|||||:|||||:|||||:|
5351 .....GACCCCTTGAAGCTCTTTAACCTGTTCGCAAT.....5319
   |||:|||||:|||||:|||||:|||||:|||||:|
182 GluPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuI 198
   |||:|||||:|||||:|||||:|||||:|||||:|
5318 .....GTCCCAATTAATGTGTGCAGTATCATTTTTCCTTTAGT 5278
   |||:|||||:|||||:|||||:|||||:|||||:|
198 ePheSerLeuGlyArgHisThrArgIleMetArgHisThrValAlaGlyS 215
   |||:|||||:|||||:|||||:|||||:|||||:|
5277 AAGATCTTATAGCAGACATACCAAGCAATAAAGCTCATGCTACGGCA 5228
   |||:|||||:|||||:|||||:|||||:|||||:|
215 eArgValProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeu 231
   |||:|||||:|||||:|||||:|||||:|||||:|
5227 GTAGGACCCAGCACAGAGTTCATGAGCAGCCATTAACCATATGACT 5178
   |||:|||||:|||||:|||||:|||||:|||||:|
232 SerPheLeuIleLeuTyPheSerHisCysMetIleLysValPheLeuSe 248
   |||:|||||:|||||:|||||:|||||:|||||:|
5177 TCATTTATCTCTCTTTTTCCTATAC.....TATATTTTC 5143
   |||:|||||:|||||:|||||:|||||:|||||:|
248 rSerLeuLysPheHisIleArgArgPheIlePheLeu.....260
   |||:|||||:|||||:|||||:|||||:|||||:|
5142 TTCTATAT.....TTGATGACCTTTAGCTATCTATGACAAATATACA 5102
   |||:|||||:|||||:|||||:|||||:|||||:|
261 .....PhePheIleLeuValIleGlyIleTyPProSerGly 272
   |||:|||||:|||||:|||||:|||||:|||||:|
5101 AGTACGTGTGAGACTTTGGAGAGATTCAGACAAATTCCTACCCCTGGGT 5052
   |||:|||||:|||||:|||||:|||||:|||||:|
273 HisSerLeuIleLeuIleLeuGlyAsnProLysLeuLysLnsAlaIat 289
   |||:|||||:|||||:|||||:|||||:|||||:|
5051 CACTCAGCTTATTTTAAATGTTTAAATATAACTGAGGAGACATTTGT 5002

```

```

289 slyspheleu 292
      :|||||
5001 CAGATGCTG 4992

seq_name: gb_pr4:AF227135
seq_documentation_block:
LOCUS      AF227135      939 bp      DNA      PRI      18-MAR-2000
DEFINITION Homo sapiens candidate taste receptor T2R9 gene, complete cds.
ACCESSION  AF227135
VERSION    AF227135.1  GI:7262616
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 939)
            Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekar, J., Ryba, N.J. and
            Zuker, C.S.
            A novel family of mammalian taste receptors
            Cell 100 (6), 693-702 (2000)
TITLE      JOORNAL
MEDLINE    20222571
REFERENCE  2 (bases 1 to 939)
            Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekar, J., Ryba, N.J. P.
            and Zuker, C.S.
            Direct Submission
            Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
            Drive, Bethesda, MD 20892, USA
FEATURES
    source
        1..939
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="12"
            /map="12p13"
            1..19
            1..939
            /note="G protein-coupled receptor; PCR-derived sequence;
            similar to an interval in BAC AC006518"
            /codon_start=1
            /product="candidate taste receptor T2R9"
            /protein_id="AAF43908.1"
            /db_xref="GI:7262617"
            /translation="MPSAIEAIVIIILAGELTIGMGFIVLNCIDMLKRRDISL
            DITSLAISRICLCVTSIDGFMLEPRGTGNSVIVNVVTPANNSLMETS
            LSTYFLKIANISHEFFERWLKLNKRWALILSLISLISVPKNDMMYHLEKVS
            HEENTWKEKVKIPGTEKQTLNGVAVPFLICISFLLSLVHTKQIRLHATG
            FRDPSTEAMRAIKAVIIFLLILIVYPVFLMSSALIPQKLVLMIGDIVTIFPS
            SHSFLINGSKLREAFKMLRFVFCFLRRKPEVP"
            <1..>939
            /product="candidate taste receptor T2R9"
    primer_bind
        248 a 187 c 186 g 318 t
        BASE COUNT 919. 939
        ORIGIN
            mRNA
            /product="candidate taste receptor T2R9"
            primer_bind
                248 a 187 c 186 g 318 t
            alignment_scores:
                Quality: 371.50      Length: 314
                Ratio: 1.905      Gaps: 9
                Percent Similarity: 62.102      Percent Identity: 32.166
            alignment_block:
                US-09-510-332-1 x AF227135
            Align seg 1/1 to: AF227135 from: 1 to: 939
            8 11eTyrPheLeuAlaValIleGlnPheLeuLeuGlyIlePheThrAs 24
            |||||:|||||
            |||||:|||||
            22 ATATATATATATTTAATGCTGGTGAATTGACCATGAGGATTTGGGAAA 71
            |||||:|||||
            24 nGlyIleIleValValValAsnGlyIleAsPheLeuIleLysHisArgLysM 41
            |||||:|||||
            72 TGGATTCATCTGTACTAGTTAATGCTGCTGCTCAAAAGAGAGATA 121

```

```

41 eTAlaProLeuAspLeuLeuLeuSerCysLeuAlaValSerArgIlePhe 57
      :|||||
122 TTTCCTGATTCATCATCATCTCTGATTCAGCTTGGCCATCTCCAGAACTGT 171
      :|||||
58 LeuGlnLeuPheLeu.....PheTyrValAsnValIleValI 70
      |||
172 CTGCTGTGTGTAATATATATAGATGCGCTCTTATTCCTGCTCTTCCAGG 221
      |||
70 ePhePheIleGlnPheIleMetCysSerAlaAsnCysAlaIleLeuLeuP 87
      :|||
222 TACATATGGCAATAGCGCTGATGATGACATTGCAATGTTGCTGGCAGAT 271
      :|||
87 heIleAsnGlnLeuGlnLeuTyrPheLeuAlaThrTyrLeuGlyValPheTyr 103
      |||||
272 TTGCCAATATTCAGATCTCTGCTTACTTCTTCCACAGATCTTCTAT 321
      :|||||
104 CysAlaValAlaSerValArgHisProLeuPheIleTyrLeuLysMe 120
      :|||||
322 TTATCAAGATAGCCAAATATATCGCACCCATTTTCTTGGCTGAAGCT 371
      :|||||
120 TArgIleSerLysLeuValProTyrPheIleLeuGlySerLeuLeuTyrV 137
      :|||||
372 AAAGATCAACAAGATCATGCTTGGCATTTCTTGGGCTCTTCTT...A 418
      :|||||
137 alSerMetIleCysValPheHisSerLysTyrAlaGlyPheMetValPro 153
      :|||||
419 TCTCTTTAATT.....ATTAGTCTTCCA 441
      :|||||
154 .....TyrPheLeuArgLysPhe...PheSerGlnAs 163
      :|||||
442 AAGATGATGATATGTGTGATATCCTTTCAAAAGTCAGTCATGAAGAANA 491
      :|||||
163 nAlaThrIleGln.....LysGlnA 170
      :|||||
492 CATTTCTTGGAATTCAAAGTACGATTAATTCAGTACTTTCANAAGCT 541
      :|||||
170 sPThrLeuAlaIleGlnIlePheSerPheValAlaGlnPheSerValPro 186
      :|||||
542 TAACCTCGTAACCTGGGGGCTGATG.....GTTCC 570
      :|||||
187 LeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyAr 203
      :|||||
571 TTTATCCTTGGCTGATTCATTTTCTGTGATCTTTCTCCCTAGTTAG 620
      :|||||
203 gHisThrArgIleMetArgAsnThrValAlaGlySerArgValProGlyA 220
      :|||||
621 ACACACCAACAGCATTCGATGCTACAGGGTCAAGAACCCAGATA 670
      :|||||
220 rGlyValaProIleSerAlaLeuLeuSerTleLeuSerPheLeuIleLeu 236
      :|||
671 CAGAGGCCCAATGAGGCCCATTAAGGCAAGTATCTTTCTGCTCCCTC 720
      :|||||
237 .....TyrPheSerHisCysMetIleLysValPheLeuSerSerle 250
      :|||||
721 CTCATCGTGTACTATCCAA.....GTCTTTCTGTATGACCTCTAGGCC 764
      :|||
250 sLysPheHisIleArgArgPheIlePheLeuPhePheIleValIleG 267
      :|||
765 TCGATTCCTCCAGGAATAATAGTATGATGATGATGATGATGATGATGAT 814
      :|||||
267 LyIleTyrProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProLys 283
      :|||||
815 TCATTTTCCCATCAAGCCATTCATCTCTTAATTTAGGGGAATAAGCAG 864
      :|||||
284 LeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCys 297
      :|||||
865 TTGAGGGAAGCTTTTCTGAAGATGTTAAGATTTGTGTGAAGCT 906

```

6  
:  
:  
:

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 12:57:24 : Search time 71 Seconds  
(without alignments)  
493.594 Million cell updates/sec

Title: US-09-510-332-1  
Perfect score: 1521  
Sequence: 1 MLESHLLIYFLAVIQFLG.....GNPKIKONAKKFLHKKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 374700 segs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:  
2: SP-bacteria:  
3: SP-fungi:  
4: SP-human:  
5: SP-invertebrate:  
6: SP-mammal:  
7: SP-mhc:  
8: SP-organelle:  
9: SP-phage:  
10: SP-plant:  
11: SP-rodent:  
12: SP-virus:  
13: SP-vertebrate:  
14: SP-unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1521	100.0	299	4 Q9NYW7	Q9NYW7 homo sapien
2	782	51.4	335	11 Q9JKT2	Q9JKT2 mus musculu
3	767	50.4	335	11 Q9JKT1	Q9JKT1 rattus norv
4	405	26.6	316	4 Q9NYW6	Q9NYW6 homo sapien
5	376	24.7	309	4 Q9NYW2	Q9NYW2 homo sapien
6	376	24.7	318	4 Q9NYW3	Q9NYW3 homo sapien
7	371.5	24.4	312	4 Q9NYW1	Q9NYW1 homo sapien
8	369	24.3	297	11 Q9JKE9	Q9JKE9 rattus norv
9	358	23.5	307	4 Q9NYW0	Q9NYW0 homo sapien
10	356	23.4	303	4 Q9NTV9	Q9NTV9 homo sapien
11	355.5	23.4	308	11 Q9JKE7	Q9JKE7 rattus norv
12	339.5	22.3	314	11 Q9JKT6	Q9JKT6 rattus norv
13	331.5	21.8	305	11 Q9JKT7	Q9JKT7 rattus norv
14	317.5	20.9	308	11 Q9JKT9	Q9JKT9 rattus norv
15	317	20.8	309	11 Q9JKT5	Q9JKT5 rattus norv
16	315.5	20.7	299	4 Q9NYW4	Q9NYW4 homo sapien
17	315.5	20.7	333	11 Q9JKT0	Q9JKT0 rattus norv
18	313	20.6	291	4 Q9NYV7	Q9NYV7 homo sapien
19	310.5	20.4	317	4 Q9NYV8	Q9NYV8 homo sapien

20	298	19.6	299	4 Q9NYW5	Q9NYW5 homo sapien
21	294	19.3	300	11 Q9JKT4	Q9JKT4 mus musculu
22	289	19.0	309	11 Q9JKT8	Q9JKT8 rattus norv
23	266.5	17.5	299	11 Q9JKT0	Q9JKT0 rattus norv
24	264.5	17.4	297	11 Q9JKT3	Q9JKT3 mus musculu
25	242.5	15.9	246	11 Q9JKA0	Q9JKA0 mus musculu
26	237.5	15.6	245	11 Q9JKA1	Q9JKA1 mus musculu
27	231.5	15.2	245	11 Q9JKA4	Q9JKA4 mus musculu
28	230.5	15.2	243	11 Q9JKA2	Q9JKA2 mus musculu
29	217.5	14.3	238	11 Q9JKA3	Q9JKA3 mus musculu
30	182	12.0	131	11 Q9JKE8	Q9JKE8 rattus norv
31	134	8.8	410	13 Q9PMU2	Q9PMU2 typhloecte
32	130.5	8.6	354	13 Q9PMU3	Q9PMU3 mus musculu
33	128.5	8.4	354	13 Q9PT44	Q9PT44 brachydanto
34	125	8.2	287	13 Q9PT44	Q9PT44 pantodon bu
35	125	8.2	352	13 Q9JBS6	Q9JBS6 labetrophe
36	125	8.2	382	13 Q9JUK8	Q9JUK8 caenorhabdi
37	124	8.2	346	13 Q9PUA9	Q9PUA9 bufo japoni
38	124	8.2	500	8 Q9ZQZ2	Q9ZQZ2 cyanidlosch
39	122.5	8.1	335	11 Q6Z853	Q6Z853 rattus norv
40	122.5	8.1	547	5 Q45965	Q45965 caenorhabdi
41	122	8.0	352	13 Q9JBS5	Q9JBS5 drosophila
42	121.5	8.0	327	11 Q70271	Q70271 rattus norv
43	121.5	8.0	341	8 Q9MD83	Q9MD83 dirosophila
44	120.5	7.9	327	11 Q70269	Q70269 rattus norv
45	120.5	7.9	342	12 Q90387	Q90387 kaposi's sa

ALIGNMENTS

RESULT	1	PRELIMINARY	PRT	299 AA.
Q9NYW7	Q9NYW7			
AC	Q9NYW7			
DT	01-OCR-2000 (TREMBLrel. 15, Created)			
DT	01-OCR-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCR-2000 (TREMBLrel. 15, Last annotation update)			
DE	CANDIDATE TASTE RECEPTOR T2RL.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Adler E., Hoon M.A., Mueller K.L., Chandrasekhar J., Ryba N.J.P.,			
RA	Zuker C.S.,			
RT	"A novel family of mammalian taste receptors."			
RL	Cell 100:693-702(2000).			
DR	EMBL: AF227129; AAF43902.1; -			
KW	Receptor.			
SEQ	SEQUENCE 299 AA; 34333 MW; CFB17DD15645392C CRC64;			
Query Match	100.08;	Score 1521;	DB 4;	Length 299;
Best Local Similarity	100.08;	Pred. NO. 2.5e-97;		
Matches 299;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 MLESHLLIYFLAVIQFLGIFTNGIIVVNGIDIKRKNAPDLILSCIAVSRIQL 60			
DB	1 MLESHLLIYFLAVIQFLGIFTNGIIVVNGIDIKRKNAPDLILSCIAVSRIQL 60			
QY	61 FTFYVNVIVIFPIEFIMCSANCAILLFTNELMLATWLGVYCKAVSVRRPFTWKM 120			
DB	61 FTFYVNVIVIFPIEFIMCSANCAILLFTNELMLATWLGVYCKAVSVRRPFTWKM 120			
QY	121 RISKLVPMMLISLIVSMICVFSKVGEMVYFLRKFSSONATQREDTLAIQIFSV 180			
DB	121 RISKLVPMMLISLIVSMICVFSKVGEMVYFLRKFSSONATQREDTLAIQIFSV 180			
QY	181 AERSVPLILFLFAVLLIFSLGHRTRQMRNTVAGSRVPGRGAPISALLSIFLLIFSH 240			
DB	181 AERSVPLILFLFAVLLIFSLGHRTRQMRNTVAGSRVPGRGAPISALLSIFLLIFSH 240			

OY 241 CMIKVFLSLKFIHRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCCQ 299  
 DB 241 CMIKVFLSLKFIHRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCCQ 299

RESULT 2  
 OYJKT2 PRELIMINARY; PRT; 335 AA.  
 AC O9JKT2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R19.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVJ;  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227149; AAF43922.1; -  
 KW Receptor  
 SQ SEQUENCE 335 AA; 37734 MW; 1399C323FC7CE645 CRC64;

Query Match 51.4%; Score 782; DB 11; Length 335;  
 Best Local Similarity 51.5%; Pred. No. 1.2e-46;  
 Matches 154; Conservative 51; Mismatches 90; Indels 4; Gaps 3;

OY 1 MESHLLIYFLAVIQFLGIFTNGIIVVNGIDLKRRMAPDLISCLAVSRIFLQ 60  
 DB 1 MMEGHMFLFLVVOFVLGVLVNGLIYVNAIDLMKKMAPDLISCLAVSRIFLQ 60  
 OY 61 FIFVYVYVFFIEFIMCSANCAILFNELEMLATWGVFCAYASVRHPLFTWLM 120  
 DB 61 CILFAQGLSFLVRLHPLFADNVTFVYITNELSLMFWLGVFCAYATIPHPLFTWLM 120  
 OY 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 DB 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 OY 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 DB 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 OY 180 VAESVPLIFLFAVILLIFSLGRHTRONRVAGSRVGRCAPISALISFLILVS 239  
 DB 180 VAESVPLIFLFAVILLIFSLGRHTRONRVAGSRVGRCAPISALISFLILVS 239  
 OY 181 --GUTLPLFLFYAVALLISLNNHSHOMR-TWVGREPSRHALVSAMLSIFLILYS 237  
 DB 181 --GUTLPLFLFYAVALLISLNNHSHOMR-TWVGREPSRHALVSAMLSIFLILYS 237  
 OY 240 HCMIKVFLSLKFIHRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 240 HCMIKVFLSLKFIHRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 238 HDMAVAVLITGCGHFGSRFAFLVIGMYPRLSHIVLILGNPKIKONAKKFLHSHKCC 296

RESULT 3  
 OYJKT2 PRELIMINARY; PRT; 335 AA.  
 AC O9JKT2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).

DR EMBL: AF227140; AAF43913.1; -  
 KW Receptor  
 SQ SEQUENCE 335 AA; 37851 MW; AB8A1CEDDE70C12C CRC64;

Query Match 50.4%; Score 767; DB 11; Length 335;  
 Best Local Similarity 51.8%; Pred. No. 1.3e-45;  
 Matches 155; Conservative 53; Mismatches 87; Indels 4; Gaps 4;

OY 1 MESHLLIYFLAVIQFLGIFTNGIIVVNGIDLKRRMAPDLISCLAVSRIFLQ 60  
 DB 1 MMEGHMFLFLVVOFVLGVLVNGLIYVNAIDLMKKMAPDLISCLAVSRIFLQ 60  
 OY 61 FIFVYVYVFFIEFIMCSANCAILFNELEMLATWGVFCAYASVRHPLFTWLM 120  
 DB 61 CILFAQGLSFLVRLHPLFADNVTFVYITNELSLMFWLGVFCAYATIPHPLFTWLM 120  
 OY 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 DB 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 OY 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 DB 121 RISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 179  
 OY 180 VAESVPLIFLFAVILLIFSLGRHTRONRVAGSRVGRCAPISALISFLILVS 239  
 DB 180 VAESVPLIFLFAVILLIFSLGRHTRONRVAGSRVGRCAPISALISFLILVS 239  
 OY 179 VGLTLPFLFYAVALLISLNNHSHOMR-TWVGREPSRHALVSAMLSIFLILYS 237  
 DB 179 VGLTLPFLFYAVALLISLNNHSHOMR-TWVGREPSRHALVSAMLSIFLILYS 237  
 OY 240 HCMIKVFLSLKFIHRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 240 HCMIKVFLSLKFIHRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 238 HDMAVAVLITGCGHFGSRFAFLVIGMYPRLSHIVLILGNPKIKONAKKFLHSHKCC 296

RESULT 4  
 OYNYW6 PRELIMINARY; PRT; 316 AA.  
 AC O9NYW6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227130; AAF43903.1; -  
 KW Receptor  
 SQ SEQUENCE 316 AA; 35914 MW; 6A7BA3C6497ABC5B CRC64;

Query Match 26.6%; Score 405; DB 4; Length 316;  
 Best Local Similarity 36.4%; Pred. No. 8.1e-21;  
 Matches 114; Conservative 43; Mismatches 114; Indels 42; Gaps 8;

OY 8 IYFLAVIQFLGIFTNGIIVVNGIDLKRRMAPDLISCLAVSRIFLQ 67  
 DB 8 IYFLAVIQFLGIFTNGIIVVNGIDLKRRMAPDLISCLAVSRIFLQ 67  
 OY 68 IYFIFEFIMCSANCAILFNELEMLATWGVFCAYASVRHPLFTWLM 119  
 DB 68 IYFIFEFIMCSANCAILFNELEMLATWGVFCAYASVRHPLFTWLM 119  
 OY 67 ---FLIESPHTHDSGITMOIYDSWFTTNLSLWATCLGVLCLKIASFSHPFTWLM 123  
 DB 67 ---FLIESPHTHDSGITMOIYDSWFTTNLSLWATCLGVLCLKIASFSHPFTWLM 123  
 OY 120 WRISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 168  
 DB 120 WRISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 168  
 OY 124 WRISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 183  
 DB 124 WRISLVPWMLIGSLVSMICVFSKYGAFVYFLRKFFSONAT-IQKEDTLAIOIFS 183  
 OY 169 EDTLAIOIFSVAESVPLIFLFAVILLIFSLGRHTRONRVAGSRVGRCAPISAL 228  
 DB 169 EDTLAIOIFSVAESVPLIFLFAVILLIFSLGRHTRONRVAGSRVGRCAPISAL 228  
 OY 184 LGTL-----WLPPLIVLASYSLLIFSLGRHTRONRVAGSRVGRCAPISAL 234  
 DB 184 LGTL-----WLPPLIVLASYSLLIFSLGRHTRONRVAGSRVGRCAPISAL 234

QY 229 SIUSF---LITYSHCKIKVFLSL-KFHRRIFLFFLIVIGYPSGSLILGNPKL 284  
 DB 235 ILLSFFFLFYFLAFIAPISFGNLPKTKAKMIGEVMT---FTPAGHSFILLIGNSKL 291  
 QY 285 KONAKKFLHSKC 297  
 DB 292 KOT---FVVMIRC 301

## RESULT 5

Q9NWM2 PRELIMINARY: PRT: 309 AA.  
 AC Q9NWM2.000 (TREMELREL. 15, Created)  
 DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RA ADLER E., HOON M.A., MUELLER K.L., CHANDRASHEKAR J., RYBA N.J.P.,  
 RA ZUKER C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL; AF227134; AAF43907.1; -  
 KW Receptor.  
 SO SEQUENCE 309 AA; 35877 MW; 5E3D9AB726A52413 CRC64;

Query Match 24.7%; Score 376; DB 4; Length 309;  
 Best Local Similarity 31.6%; Pred. No. 7.7e-19;  
 Matches 101; Conservative 56; Mismatches 101; Indels 62; Gaps 10;

QY 8 IYELLAVIOFLGIFNGIIVVNGIDLKHKMAPLDLLSCLAVSRIFLOLFIYVNV 67  
 DB 8 IYILLAGEFLTIGNGIIVVNGIDLKHKMAPLDLLSCLAVSRIFLOLFIYVNV 66  
 QY 68 IYVFFIFMGSCANCAILL-----FINELEMLATWLGVCYCAKVASVRHPLFTWLMKRI 122  
 DB 67 IYVILNDVYVTKKQOIVITFEFTFANYLMWMTTCTCINAVYFLKIKSSSHPLFTWLMKRI 126  
 QY 123 SKVPMMLISL--LYVSMICV-----FHS-----KXAGEVVPFLKFRS 161  
 DB 127 DMVYHMLILGFAISLIVSLIAIVLSCDYRFHAIKHKRNITEMPHVSKIPF----- 180  
 QY 162 ONATIQEDTLAIOISFVAFESVPLILFEAVILLIFSGRHTROMRNTVAGSRVPRG 221  
 DB 181 -----EPLTLFNLFI-----VPFVLSISFPLVRSIMRHTKQIKLYATGSRDPSTE 228  
 QY 222 APISALLSILSFLYSHCKIKVFLSLKFIHRRIFL-----FFLIVIGIYPSG 272  
 DB 229 VHVRAIKMTSFIFFFLIX-----YISSI---LMTFSYIMTKYKLAVEFGHAIILYPLG 280  
 QY 273 HSLILILGNPKLONAKKFL 292  
 DB 281 HSLILILGNPKLONAKKFL 300

## RESULT 6

Q9NWM3 PRELIMINARY: PRT: 318 AA.  
 AC Q9NWM3.000 (TREMELREL. 15, Created)  
 DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA ADLER E., HOON M.A., MUELLER K.L., CHANDRASHEKAR J., RYBA N.J.P.,  
 RA ZUKER C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL; AF227133; AAF43906.1; -  
 KW Receptor.  
 SO SEQUENCE 318 AA; 36549 MW; F187ADB2D8274B8A CRC64;

Query Match 24.7%; Score 376; DB 4; Length 318;  
 Best Local Similarity 32.7%; Pred. No. 7.9e-19;  
 Matches 96; Conservative 60; Mismatches 120; Indels 18; Gaps 5;

QY 12 LAVIOFLGIFNGIIVVNGIDLKHKMAPLDLLSCLAVSRIFLOLFIYVNVYIF 71  
 DB 12 LAVGEFSVIGILGNAFIGLVCMQMKRKRKIASIDLITSLAISRLCLLCVILDCFTVL 71  
 QY 72 FIEFIMSCANCAILLFI-----NELEMLATWLGVCYCAKVASVRHPLFTWLMKRI 127  
 DB 72 YPDVYATGKEMRIIDFEWTLNHLSTIFATCLSIYFFKIGNFHFPLMKMRIDRIVIS 131  
 QY 128 WILGSLIYVSMICVFSKXA-----GFWVPYLRKFFSONATIQEDTLAIOISF 179  
 DB 132 WILGCV-----VLSVFLSPATENLNDFFRCVAKRKTMLTSCRVNKHASTKFLFN 187  
 QY 180 VAEFSVPLILFEAVILLITSLGRHTROMRNTVAGSRVPGAPISALLSILSFLIYFS 239  
 DB 188 LATL-LPFCVCLMSFLLILSLRHRIRMOISATGCDPSTEAVHRAKVAISFLLFLFA 246  
 QY 240 HCMIKVFLSLKFIHRRIFLFFLIVIG-IYPSGSHLILILGNPKLONAKKFL 292  
 DB 247 YLISFLITSSYFMPPEFLAVIFGESTALITYPSSSHFLILGNPKLONAKKFL 300

## RESULT 7

Q9NWM1 PRELIMINARY: PRT: 312 AA.  
 AC Q9NWM1.000 (TREMELREL. 15, Created)  
 DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R9.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RA ADLER E., HOON M.A., MUELLER K.L., CHANDRASHEKAR J., RYBA N.J.P.,  
 RA ZUKER C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL; AF227135; AAF43908.1; -  
 KW Receptor.  
 SO SEQUENCE 312 AA; 35611 MW; CCE66F1E31051F49 CRC64;

Query Match 24.4%; Score 371.5; DB 4; Length 312;  
 Best Local Similarity 32.1%; Pred. No. 1.6e-18;  
 Matches 100; Conservative 60; Mismatches 113; Indels 39; Gaps 8;

QY 8 IYELLAVIOFLGIFNGIIVVNGIDLKHKMAPLDLLSCLAVSRIFLOLFI---F 63  
 DB 8 IYILLAGEFLTIGNGIIVVNGIDLKHKMAPLDLLSCLAVSRIFLOLFI---F 67  
 QY 64 YVNVYVFFIEFIMSCANCAILLFINELEMLATWLGVCYCAKVASVRHPLFTWLMKRI 123  
 DB 68 FMLFPCTGNSVLYVNVVWTFANNSLMPFSCLSIFFLTKIANISHPFFFWLTKIN 127  
 QY 124 KIVPMMLISLIVVSMICVFSKXAGEVMP-----YFLKRF-FSONATIO----- 167

DB 128 KVMALILGSL-LSL-----ISVKNDDMMVHLFKVSHENITWKFVKISPGT 177  
 QY 168 -KEDFLAIQIFSVAFESVPLIFLFAVLLIFESGCRTRORNTVASSRPGRCAPISA 226  
 DE 178 FQGLTINLGV-----VFPLICISFLLFSVLRHTKQIRLHATFROPSTEAHRA 230  
 QY 227 LLSISFLIYSHC-MIKVPLSLKFIHRIREFLFIYIGIYPSGSHLLILGNPKLK 285  
 DB 231 IKAIVIFLLLIYVYVFLVMTSSALIPQGLVLMIGDIVIYIPSSHSFTLLNGNSKLR 290  
 QY 286 QNAKFFLHSC 297  
 DB 291 EAFKMLRFVK 302

RESULT 8  
 QJKE9  
 ID 09JKE9 PRELIMINARY: PRT: 297 AA.  
 AC 09JKE9  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE TASTE RECEPTOR RT2R6 (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN NCBI\_Taxid=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF240766; AAF43304.1; -.  
 KW Receptor.  
 FT NON\_TER  
 SQ SEQUENCE 297 AA; 34107 MW; 873D0440C2863E25 CRC64;

Query Match 24.38; Score 369; DB 11; Length 297;  
 Best Local Similarity 34.18; Pred. No. 2, 2e-18;  
 Matches 104; Conservative 51; Mismatches 106; Indels 44; Gaps 9;

QY 18 LIGFTNGIIVVNGIDILIKHRKMAPDLLSCLAISRIFLOFIYVNYVIFEFIEFM 77  
 DB 3 LVGLGNALFALVNFMMKMKRTAIDLILSLMSRILCCLILDLVQYDPYV 62  
 QY 78 CSANCAILFT---NELEMLATWLGVCYCAVASVRHPLFIWLMKRIKSLVPMILGS 133  
 DB 63 RGMKMRIDDEFMTLNLHSLVWFATCLISFYEFKIANEFHPLFMKIRIDKILRFLAC 122  
 QY 134 LLYVMICVFSKYGFMVPI-----FLRKFFSO--NATIQ-----KEDTL 172  
 DB 123 L--LILSLC-----FSLPTVENLADDFRCYKTKERINSTLRCKLNAGAVSVNIN 172  
 QY 173 AIQIFSVVAFSPPLIFLFAVLLIFSLGRHTRORNTVASSRPGRCAPISALISLS 232  
 DB 173 LVMLF-----PFSVSLVSF-----LLILSLRHTRQMLVNTGVNDPSTAHAKATAVIS 224  
 QY 233 FLIYFSHCMIKVFSLKFIHRIREFLFIYIGIYPSGSHLLILGNPKLONAKKF 291  
 DB 225 FLVFIYVCLAFIANTISYFMPESLAVIMGELIALIYIPSSHSFTLLNGNSKLR 284  
 QY 292 LHSK 296  
 DB 285 LCRVK 289

RESULT 9  
 ID 09NIV9 PRELIMINARY: PRT: 307 AA.  
 AC 09NIV9  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.  
 RN NCBI\_Taxid=9606;  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227136; AAF43909.1; -.  
 KW Receptor.  
 SQ SEQUENCE 307 AA; 35365 MW; 058DD69A55AE71C6 CRC64;

Query Match 23.58; Score 358; DB 4; Length 307;  
 Best Local Similarity 31.08; Pred. No. 1, 3e-17;  
 Matches 96; Conservative 68; Mismatches 110; Indels 36; Gaps 12;

QY 8 IYFLAVIOFLGIFTNGIIVVNGIDILIKHRKMAPDLLSCLAISRIFLOFIYVNV 67  
 DB 8 IFIFVVSSEVFGVLNGFGLVNCIDAKN-KLSYIGILGLAISRIFL-IWIIIDG 65  
 QY 68 IVFFIEFIMCSANCAILFI-----NELEMLATWLGVCYCAKVASVRHPLFIWLM 120  
 DB 66 FIOIESPNIVASGN--LIIYISYFVWIGQSSMPATSLISFYFKINFSYIFLMKS 123  
 QY 121 RISKLVPMMILGSLIYVSMICVFSKYGFMVPIFLRKFFSONATIOKEDTL-AIQIS- 178  
 DB 124 RINMVLPMIV--FLILSLNF-----AYAIKINDYK--KNDIVWDLNMTKS 169  
 QY 179 --FVAEFSVPL-LIFLFAV-----LLIFSLGRHTRORNTVASSRPGRCAPISALLSI 230  
 DB 170 EYFIQIILNLNGVIFEFILSLITCIFIILSRHNRQMSVNTGLDSNTAHVAKMKVL 229  
 QY 231 LSFLLIYFSHCMIKVFSLKFIHRIREFLFIYIGIYPSGSHLLILGNPKLONAK 289  
 DB 230 ISFILLFIYFGMAIEISCFVRENKLLIMFGMTTITVFMGHSFTLLNGNSKLR 289  
 QY 290 KFLHSCQ 299  
 DB 290 RVLDQKCE 299

RESULT 10  
 ID 09NIV9 PRELIMINARY: PRT: 303 AA.  
 AC 09NIV9  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R13.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.  
 RN NCBI\_Taxid=9606;  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227137; AAF43910.1; -.  
 KW Receptor.  
 SQ SEQUENCE 303 AA; 35118 MW; 756BF5382E36BF5 CRC64;

Query Match 23.48; Score 356; DB 4; Length 303;  
 Best Local Similarity 31.98; Pred. No. 1, 8e-17;  
 Matches 97; Conservative 69; Mismatches 116; Indels 22; Gaps 11;







**THIS PAGE BLANK** (US)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 12:58:00 ; Search time 71.14 seconds  
(Without alignments)  
135.731 Million cell updates/sec

Title: US-09-510-332-1  
Perfect score: 1521  
Sequence: 1 MESHLLITFLAVIQFLG.....GNPKLNQAKKFLHSHKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues  
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	8.6	333	1 NU2M_APLI	P34849 apis mellif
2	130.5	8.6	348	1 OPSD_RABIT	P49912 oryctolagus
3	129.5	8.5	348	1 OPSD_SHEEP	P02700 ovis aries
4	129.5	8.5	354	1 OPSD_RAJER	P79863 raja eriac
5	128.5	8.4	348	1 OPSD_CRIGR	P28681 cricetus
6	128.5	8.4	348	1 OPSD_RAT	P51489 rattus norv
7	128.5	8.4	354	1 OPSD_BUFB	P56514 buto buto
8	125.5	8.3	348	1 OPSD_MESBI	P62798 mesopiodon
9	125	8.2	348	1 OPSD_TURTR	P62798 tursiops tr
10	123.5	8.1	312	1 OZCL_HUMAN	P08100 homo sapien
11	123.5	8.1	348	1 OPSD_HUMAN	P02791 delphinus d
12	123	8.1	348	1 OPSD_DELDE	P01493 homo sapien
13	122.5	8.1	346	1 GP41_HUMAN	P62794 phoca vitul
14	122.5	8.1	348	1 OPSD_PHOVI	P18766 sus scrofa
15	122.5	8.0	274	1 OPSD_PIG	P29667 drosophila
16	121.5	8.0	348	1 OPSD_DROMA	P29667 drosophila
17	121.5	8.0	348	1 OPSD_MACFA	P15409 mus musculu
18	121.5	8.0	348	1 OPSD_MOUSE	P62796 trichechus
19	121.5	8.0	348	1 OPSD_PRIMA	P62792 globicephal
20	121	8.0	348	1 OPSD_GLOME	P51471 anolis caro
21	121	8.0	355	1 OPSB_ANOCA	O51750 borrelia bu
22	121	8.0	512	1 MVIN_BORBU	O98146 kaposi's sa
23	120.5	7.9	342	1 VG74_KSHV	O15529 homo sapien
24	120.5	7.9	346	1 GP42_HUMAN	P32308 canis famli
25	120.5	7.9	348	1 OPSD_CANFA	O92024 anguilla an
26	120.5	7.9	352	1 OPSD_ANGAN	O93459 scyllorhinu
27	120.5	7.9	354	1 OPSD_SCYCA	P32311 carassius a
28	120	7.9	349	1 OPSG_CARAU	P22671 lampetra ja
29	119.5	7.9	353	1 GTR2_HUMAN	P11168 homo sapien
30	118.5	7.8	524	1 MLIQ_XENLA	P49229 xenopus lae
31	118	7.8	420	1 OPSD_PHOGR	O62795 phoca groen
32	117.5	7.7	348	1 OPSD_PHOGR	P12336 rattus norv
33	117.5	7.7	522	1 GTR2_RAT	

34	117	7.7	354	1 OPSD_GALML	O93441 galeus mela
35	116.5	7.7	341	1 NU2M_DROME	P03896 drosophila
36	116.5	7.7	354	1 OPSD_ZEUPA	O42604 zeus faber
37	116.5	7.7	359	1 AC2R_BOVIN	P25104 bos taurus
38	116	7.6	353	1 OPSD_DIPVU	O99404 diploodus vu
39	115.5	7.6	348	1 OPSD_BOVIN	P02699 bos taurus
40	115.5	7.6	352	1 OPSD_ANOCA	P41591 anolis caro
41	115.5	7.6	355	1 OPSG_CHICK	P28683 gallus gall
42	114.5	7.5	353	1 OPSD_PETMA	O98980 petromyzon
43	114.5	7.5	354	1 OPSD_XENLA	P29403 xenopus lae
44	114.5	7.5	362	1 RODI_RAT	O89039 rattus norv
45	114	7.5	289	1 OPSD_COTIN	O42330 cottocomeph

ALIGNMENTS

RESULT ID	1	NU2M_APLI	STANDARD	PRT	333 AA
AC	P34849	P92496	P92497	P92502	P92886; P92907; Q33785; Q33786;
AC	Q33787	Q33788	Q36885		
DT	01-FEB-1994	(Rel. 28, Created)			
DT	01-FEB-1994	(Rel. 28, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	NADH-UBIQUINONE-OXIDOREDUCTASE CHAIN 2 (BC 1.6.5.3).				
GN	N2.				
OS	Apis mellifera ligustica (Common honeybee).				
OG	Mitochondrion.				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;				
OC	Apoidea; Apidae; Apis.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=THORAX;				
RX	MEDLINE=93114603; PubMed=8417993;				
RA	Crozier R.H., Crozier Y.C.;				
RT	"The mitochondrial genome of the honeybee Apis mellifera: complete				
RT	sequence and genome organization.";				
RL	Genetics 133:97-117(1993).				
RN	[2]				
RP	SEQUENCE OF 1-211 FROM N.A.				
RC	STRAIN=VARIOUS;				
RX	MEDLINE=96364025; PubMed=8744768;				
RA	Arias M.C., Sheppard W.S.;				
RT	"Molecular phylogenetics of honey bee subspecies (Apis mellifera L.)				
RT	inferred from mitochondrial DNA sequence.";				
RL	Mol. Phylogenet. Evol. 5:557-566(1996).				
RN	[3]				
RP	SEQUENCE OF 60-307 FROM N.A.				
RC	STRAIN=HAPLOTYPES 1 TO 11;				
RA	Koulianos S., Crozier R.H.;				
RT	Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.				
CC	- CARBOXYLACTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.				
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL				
CC	INNER MEMBRANE.				
CC	- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC					
DR	EMBL: 106178; AAB96798.1; -				
DR	EMBL: 035743; AAB38223.1; -				
DR	EMBL: 035744; AAB38224.1; -				
DR	EMBL: 035745; AAB38225.1; -				
DR	EMBL: 035746; AAB38226.1; -				
DR	EMBL: 035747; AAB38227.1; -				
DR	EMBL: 035748; AAD12744.1; -				



TRANSMEM 285 309 7 (POTENTIAL).  
 DOMAIN 310 348 CYTOPLASMIC.  
 MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
 CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
 BINDING 296 296 RETINAL CHROMOPHORE.  
 BINDING 322 322 PALMITATE (BY SIMILARITY).  
 LIPID 323 323 PALMITATE (BY SIMILARITY).  
 LIPID 323 323 BY SIMILARITY.  
 DISULFID 110 187 BY SIMILARITY.  
 MOD\_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).  
 SEQUENCE 348 AA; 38994 MW; 09405AB5EA2E1E4 CRC64;

Query Match 8.6%; Score 130.5; DB 1; Length 348;  
 Best Local Similarity 21.8%; Pred. No. 0.02;  
 Matches 70; Conservative 64; Mismatches 128; Indels 59; Gaps 16;

1 MESHLLIYFLAVIOPLGIFNGIIVVNGIDLIKRRKM-APDLLSCLAVSRIFLQ 59  
 MIAAYM---FLIVLGFPINFLVTV-----QHKRLTPUNYILMLAVADLEMV 87

60 LFIPIYVAVIPIEFIFMCSANCAILFI---NELELM---LATWLGVCARVASVR 111  
 LGFTTTLTSLHGYFVPGTCNVEGFATLGGELALMSLVLAIERVYVCKPKMSNR 147

88 LGGFTTTLTSLHGYFVPGTCNVEGFATLGGELALMSLVLAIERVYVCKPKMSNR 147

112 ---HPL---FIMLKMRISKLVPMILGSLLYV---SMICVFSKYAGFMPYFLKFFS 161

148 FENNAIMKVAFTWI-MALACAP-PLVGMSRYIPEGQSCGIDY-----YTLKPEVN 199

162 QNATIQEDTLAIQIFSEVAESVLLIFLPAVLLIFSLGHTQMNIVAGSRVPRG 221

200 NESFV-----LYMEFVHFTIPIILIFFCYGLVFEVKEAAQODS-ATYKAEKE 249

222 APISALLSTSLIFLYESHCMKVFSLKFIHRIPIF---FFLIVIGVPSGSHLL 277

250 YVRVAVIYALICWPAVPAVAFIYFIHQSGNFGIIFETIAFAFKSSSIY---NPVIY 306

278 ILGNPKLONAKKFLHSKCC 298

307 IMNKKPF---RNCMLTJICC 323

RESULT 3  
 OPSD\_SHEEP STANDARD: PRT; 348 AA.  
 ID OPSD\_SHEEP PRT; 348 AA.  
 AC P02700;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 RP SEQUENCE.  
 RA Pappin D.J.C.; Elipoulos E.; Brett M.; Findlay J.B.C.;  
 RT "A structural model for ovine rhodopsin."  
 RL Int. J. Biol. Macromol. 6:73-76(1984).  
 RN [2]  
 RP SEQUENCE OF 1-111 AND 144-239.  
 RA MEDLINE=83282605; Pubmed=6224479;  
 RX Brett M.; Findlay J.B.C.;  
 RT "Isolation and characterization of the CNBR peptides from the  
 proteolytically derived N-terminal fragment of ovine opsin."  
 RL Biochem. J. 211:661-670(1983).  
 RN [3]  
 RP SEQUENCE OF 240-348.  
 RX MEDLINE=82013638; Pubmed=7278988;  
 RA Findlay J.B.C.; Brett M.; Pappin D.J.C.;  
 RT "Primary structure of C-terminal functional sites in ovine  
 rhodopsin.";

Nature 293:314-316(1981).  
 (4)  
 RP RETINAL BINDING SITE.  
 RX MEDLINE=84178280; Pubmed=6370231;  
 RA Pappin D.J.C.; Findlay J.B.C.;  
 RT "Sequence variability in the retinal-attachment domain of mammalian  
 rhodopsins."  
 RL Biochem. J. 217:605-613(1984).  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
 LINKED TO CIS-RETINAL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
 VISION IN DIM LIGHT.  
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
 BE PHOSPHORYLATED.  
 CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 495 NM.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC OPSIN SUBFAMILY.  
 PIR: A30407; OOSH.  
 GCRDB: GCR\_0194;  
 DR INTERPRO: IPR000276;  
 DR INTERPRO: IPR000732;  
 DR INTERPRO: IPR001760;  
 DR PFAM: PF00001; 7tm\_1; 1.  
 DR PRINTS: PR00237; GPCRHHODOPSIN.  
 DR PRINTS: PR00238; OPSIN.  
 DR PRINTS: PR00579; RHODOPSIN.  
 DR PROSITE: PS00237; G-PROTEIN\_RECEP\_FL\_1; 1.  
 DR PROSITE: PS00262; G-PROTEIN\_RECEP\_FL\_2; 1.  
 DR PROSITE: PS00238; OPSIN; 1.  
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;  
 KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.  
 FT DOMAIN 1 36  
 FT TRANSMEM 37 61  
 FT DOMAIN 62 73  
 FT TRANSMEM 74 98  
 FT DOMAIN 99 113  
 FT TRANSMEM 114 133  
 FT DOMAIN 134 152  
 FT TRANSMEM 153 176  
 FT DOMAIN 177 202  
 FT TRANSMEM 203 230  
 FT DOMAIN 231 252  
 FT TRANSMEM 253 276  
 FT DOMAIN 277 284  
 FT TRANSMEM 285 309  
 FT DOMAIN 310 348  
 FT CARBOHYD 2 2  
 FT CARBOHYD 15 15  
 FT BINDING 296 296  
 FT LIPID 322 322  
 FT LIPID 323 323  
 FT DISULFID 110 187  
 FT MOD\_RES 343 343  
 FT SEQUENCE 348 AA; 38991 MW; AAFD6F0DA8BAE5 CRC64;

Query Match 8.5%; Score 129.5; DB 1; Length 348;  
 Best Local Similarity 21.1%; Pred. No. 0.023;  
 Matches 69; Conservative 62; Mismatches 125; Indels 71; Gaps 16;

1 MESHLLIYFLAVIOPLGIFNGIIVVNGIDLIKRRKM-APDLLSCLAVSRIFLQ 59  
 MIAAYM---FLIVLGFPINFLVTV-----QHKRLTPUNYILMLAVADLEMV 87

39 MIAAYM---FLIVLGFPINFLVTV-----QHKRLTPUNYILMLAVADLEMV 87

60 LFIPIYVAVIPIEFIFMCSANCAILFI---NELELM---LATWLGVCARVASVR 111  
 LGFTTTLTSLHGYFVPGTCNVEGFATLGGELALMSLVLAIERVYVCKPKMSNR 147

88 FGGFTTTLTSLHGYFVPGTCNVEGFATLGGELALMSLVLAIERVYVCKPKMSNR 147

112 ---HPL---FIMLKMRISKLVPMILGSLLYV---SMICVFSKYAGFMPYFLR--- 157

148 FENNAIMKVAFTWIMALACAPPLVGM-----SHY-----IPGQSCSC 187

QY 158 --KFSQNTICKEDTLAIOIFSVFAESVPLIFLFAVLILFSLGHTROMNTVAGS 215  
 DB 188 GALYFLKKEINNE---SVIYIMFVHSIPILVIFPCYGLVFTVKEMAAQOQES-RTT 243  
 QY 216 RYRGKAPISALLSLIFLIFSHCMIKVLSLKKFHIRREIFL-FFLLVIGIYPS 271  
 DB 244 QKAEKVTMRVIMVIMVIAFLICMLPYAGVAFYFTFGSDFCPIFWTIPAFKSSSV-- 301  
 QY 272 GHSLLILGNPKLONAKFELHSCKC 298  
 DB 302 -NPVIYIMNKKF-----RNCMLTLTLC 323

RESULT 4  
 OPSD\_RAJER STANDARD; PRT: 354 AA.  
 AC P79863;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Raja erinacea (Little skate).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Squalae; Hypnosquales; Pristiogadidae; Batoidae;  
 OC Rajiformes; Rajidae; Raja.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-RETINA;  
 RX MEDLINE-97398145; PubMed-9256070;  
 RA O'Brien J., Rippes H., Al-Ubaidi M.R.;  
 RL "Molecular cloning of a rod opsin cDNA from the skate retina.";  
 CC Gene 193:141-150(1997).  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 CC MEDiate VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
 CC LINKED TO CIS-RETINAL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
 CC VISION IN DIM LIGHT.  
 CC -1- PPM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
 CC BE PHOSPHORYLATED.  
 CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 501 NM.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC OPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: U01514; AAC60251.1; -;  
 CC GCRDB: GCR\_1485; -;  
 CC DR INTERPRO: IPR000276; -;  
 CC DR INTERPRO: IPR000732; -;  
 CC DR INTERPRO: IPR001760; -;  
 CC DR PFAM: PF00001; 7tm.1; 1.  
 CC DR PRINTS: PRO00237; GPCRHOOPS.  
 CC DR PRINTS: PRO0238; OPSIN.  
 CC DR PRINTS: PRO0579; RHODOPSIN.  
 CC DR PROSITE: PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
 CC DR PROSITE: PS0262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 CC DR PROSITE: PS00238; OPSIN; 1.  
 CC KW Photoreceptor. Retinal protein. Transmembrane; Glycoprotein; Vision;  
 CC Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.  
 CC FT DOMAIN 1 36  
 CC FT TRANSMEM 37 61  
 CC FT DOMAIN 62 73  
 CC FT TRANSMEM 74 98  
 CC FT DOMAIN 99 113  
 CC EXTRACELLULAR.  
 CC 1 (POTENTIAL).  
 CC 2 (POTENTIAL).  
 CC EXTRACELLULAR.

FT TRANSMEM 114 133  
 FT DOMAIN 134 152  
 FT TRANSMEM 153 176  
 FT DOMAIN 177 202  
 FT TRANSMEM 203 230  
 FT DOMAIN 231 252  
 FT TRANSMEM 253 276  
 FT DOMAIN 277 284  
 FT TRANSMEM 285 309  
 FT DOMAIN 310 354  
 FT CARBOHYD 2 2  
 FT CARBOHYD 15 15  
 FT BINDING 296 296  
 FT LIPID 322 322  
 FT DISULFID 110 187  
 SQ SEQUENCE 354 AA; 39697 MW; 1ACBB918ADA214C2 CRC64;

Query Match 8.5%; Score 129.5; DB 1; Length 354;  
 Best Local Similarity 22.2%; Pred. No. 0.023; Indels 83; Gaps 18;  
 Matches 70; Conservative 56; Mismatches 107;

QY 8 IYFLAVIOELGIFNGIIVVNGIDLIKRRK-APDLISCLAVSRIFQLFFIYVN 66  
 DB 44 MFPLI-----LTGLPVNPLFLVY-----IOHKKLRQPLVIMILNLAVSDLFVFGFTTT 94  
 QY 67 VIYIFIEPIKMCANCAILFT-----NELEMLATLVG-----FYCAKVASV-----HPL 114  
 DB 95 IITSMNGYFLFPGAGNFEQFEFATLGEVGLMCLVLAIERVYVCKPMANFRFGSOHAI 154  
 QY 115 ----FTLMKRIS----KIVPMIIGSLIYVSMICVFHSKYVGFMPYELR-----KFS 161  
 DB 155 IGWVFWI-MALSCAPPLVG-----SR-----IPEGLQSCGDYIT 193  
 QY 162 QNATICKEDTLAIOIFSVFAESVPLIFLFAVLILFSLGHTROMNTVAGSRVPRG 221  
 DB 194 MKREVNE---SVIYIMFVHSIPILVIFPCYGLVFTVKEMAAQOQESSTORAE-RE 249  
 QY 222 APISALLSLIFLIFSHCMIKVLSLKKFHIRREIFL-FFLLVIGIYPSG- 272  
 DB 250 VYRMVIMVIAFLICMLPYAGVAFYFTFGSDFCPIFWTIPAFKSSSV 298  
 QY 273 ---HSLILGNPKK 285  
 DB 299 AVYNPLIYIMNKKFR 314

RESULT 5  
 OPSD\_CRIGR STANDARD; PRT: 348 AA.  
 AC P28681;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Cricetus.  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 RC TISSUE-OVARY;  
 RX MEDLINE-92219256; PubMed-1560457;  
 RA Gale J.M., Tobey R.A., D'Anna A.;  
 RL "Localization and DNA sequence of a replication origin in the  
 RT rhodopsin gene locus of Chinese hamster cells.";  
 RL J. Mol. Biol. 224:343-358(1992).  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 CC MEDiate VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
 CC LINKED TO CIS-RETINAL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES

OY 162 ONMATIOEDTLAIOIFESVAEFSVPLLIFLPAVALLIFSLSGHTQMRNTVAGSVPBG 221  
:  
:  
:  
D 200 NESPV-----IYAFVFHPIPLIVIFECYGJLVTFKBAQAODSES-ATTQAEKE 249  
:  
:  
:  
OY 222 APIALLSIISFLIVESHCMIKVPLSSSKFHIRRFILF----FFLIYGIYPSGSHLIT 277  
:  
:  
:  
D 250 VTRVMILMVAVFEFLCWPIYAGVAFIYITHOGSGNFGCIPFMILPAFAFKSSSIY---NPVIY 306  
:  
:  
:  
OY 278 ILGNPKLKONAKKKELHSKC 298  
:  
:  
:  
D 307 IMANKQF-----RNCMLTFLCC 323  
:  
:  
:  
  
RESULT 6  
OPSD\_RAT            STANDARD;         PRT;      348 AA.  
ID   OPSD\_RAT            STANDARD;         PRT;      348 AA.  
AC   PS1489;  
DT   01-OCT-1996 (Rel. 34, Created)  
DT   01-OCT-1996 (Rel. 34, last sequence update)  
DT   15-JUL-1999 (Rel. 38, last annotation update)  
DE   RHODOPSIN.  
GN   RHO.  
OS   Rattus norvegicus (Rat).  
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN   (1)  
RP   SEQUENCE FROM N.A.  
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=RETINAL ROD CELL;  
RX   MEDLINE=95383121; PubMed=7654522;  
RA   Barnstable C.J., Morabito M.A.;  
RT   "Isolation and coding sequence of the rat rod opsin gene."  
RL   J Mol. Neurosci. 5:207-209(1994).  
RN   (2)  
RS   SEQUENCE FROM N.A.  
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=RETINA;  
RA   Huber A., Baker B.B., Sander P., Gerton G., Paulsen R.,  
Williams T.P.;  
RL   Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC   -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
LINKED TO CIS-RETINAL.  
CC   -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC   -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
VISION IN DIM LIGHT.  
CC   -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
BE PHOSPHORYLATED.  
CC   -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 495 NM.  
CC   -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC   OPSIN SUBFAMILY.  
CC   -----  
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC   the European Bioinformatics Institute. There are no restrictions on its  
CC   use by non-profit institutions as long as its content is in no way  
CC   modified and this statement is not removed. Usage by and for commercial  
CC   entities requires a license agreement (See http://www.isb.ch/announce/  
or send an email to license@isb-sib.ch).  
CC   -----  
CC   EMBL: U22180; AAA84439.1; -  
DR   EMBL: Z46957; CAAB7081.1; -  
DR   GCADB: GCR\_1458; -  
DR   INTERPRO: IPR000276; -  
DR   INTERPRO: IPR000732; -  
DR   INTERPRO: IPR001760; -  
DR   PFAM: PF00001; 7tm\_1; 1.  
DR   PRINTS: PR00237; GPCR\_RHODOPSIN.  
DR   PRINTS: PR00238; OPSIN.  
DR   PRINTS: PR00579; RHODOPSIN.  
DR   PROSITE: PS00237; G\_PROTEIN\_RECPT\_FL\_1; 1.  
DR   PROSITE: PS50262; G\_PROTEIN\_RECPT\_FL\_2; 1.  
DR   PROSITE: PS00238; OPSIN; 1.  
KW   Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;

Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;  
 Acetylation.  
 KW DOMAIN 1 36 EXTRACELLULAR.  
 FT TRANSMEM 37 61 1 (POTENTIAL).  
 FT TRANSMEM 62 73 CYTOPLASMIC.  
 FT TRANSMEM 74 98 2 (POTENTIAL).  
 FT TRANSMEM 99 113 EXTRACELLULAR.  
 FT TRANSMEM 114 133 3 (POTENTIAL).  
 FT TRANSMEM 134 152 CYTOPLASMIC.  
 FT TRANSMEM 153 176 4 (POTENTIAL).  
 FT TRANSMEM 177 202 EXTRACELLULAR.  
 FT TRANSMEM 203 230 5 (POTENTIAL).  
 FT TRANSMEM 231 252 CYTOPLASMIC.  
 FT TRANSMEM 253 276 6 (POTENTIAL).  
 FT TRANSMEM 277 284 EXTRACELLULAR.  
 FT TRANSMEM 285 309 7 (POTENTIAL).  
 FT TRANSMEM 310 348 CYTOPLASMIC.  
 FT MOD RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
 FT BINDING 15 15 RETINAL CHROMOPHORE.  
 FT LIPID 296 296 PALMITATE (BY SIMILARITY).  
 FT LIPID 322 322 PALMITATE (BY SIMILARITY).  
 FT DISULFID 323 323 BY SIMILARITY.  
 FT MOD RES 110 187 BY SIMILARITY.  
 FT CONFLICT 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).  
 SQ SEQUENCE 348 AA: 39042 MW: DA0F3F90C30984BC CRC64;  
 S -> T (IN REF. 2).

Query Match 8.4%; Score 128.5; DB 1; Length 348;  
 Best Local Similarity 20.5%; Pred. No. 0.027; Indels 71; Gaps 16;  
 Matches 67; Conservative 64; Mismatches 125;

QY 1 MESHLLIYVLLAVIOFLGIFNGIIVVNGIDILKHKM-APDLLSCLAVSRIFLO 59  
 DB 39 MLAAYM---FLIVLGFPIINFLVTV-----QHKLRPLIILLNLAVALEPW 87  
 QY 60 LTFYVYVYVIFIEFIEFMCANAILLFI---NEELM---LATWLGFFYCAKVASVR 111  
 DB 88 FGGFTTLLYSLGIVFVGTCNLGEPATLGGELGMSLVLAERYVAVVCKPMSNR 147  
 QY 112 ----HPL----FTW----LKNRSLKLVPMILSLVSMICVFSKYGEMVPIELR--- 157  
 DB 148 FGENHAIMGVAFWVAALCAAPLVGM-----SRV-----IPEGKQSC 187  
 QY 158 --KEFSQNTIQKEDLAIQISFVAESVPLILFPAVLLIFSLGRTRQMTVAGS 215  
 DB 188 GIDYTLKPEVNE--SFVYIMFVNHFTIPMIVIFFCYQGLVTVKEXAAQOQS-ATT 243  
 QY 216 RVPGRGAPISALSLIFLVSFKMIVFLSLKPHIRRIPL-----FTLVIGITPS 271  
 DB 244 QNAKEVETMVIIMVIFFLICMLPYASVAMFTFHQGSNFGPIFTLPAFAKTVASTY-- 301  
 QY 272 GHSLLIILGNPKLONAKKFLHRSKC 298  
 DB 302 -NPITYIMNMKQF---RNCMLTSLQC 323

RESULT 7  
 OPSD\_BUFBU STANDARD; PRT: 354 AA.  
 AC P56514.  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Bufo bufo (European toad).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae;  
 CC Bufo.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE-98288405; PubMed-9533857;  
 RA Fyhrquist N., Donner K., Hargrave P.A., McDowell J.H., Popp M.P.,  
 RA Smith W.C.,  
 RT "Rhodopsins from three frog and toad species: sequences and  
 RT functional comparisons."  
 RL Exp. Eye Res. 66:295-305(1998).  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
 CC LINKED TO CIS-RETINAL. THIS RECEPTOR IS COUPLED TO THE ACTIVATION  
 CC OF PHOSPHOLIPASE C.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
 CC BE PHOSPHORYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC OPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: 059931; AAB93704.1; .  
 DR GCRDB: GCR\_2464; .  
 DR INTERPRO: IPR000276; .  
 DR INTERPRO: IPR000732; .  
 DR INTERPRO: IPR001760; .  
 DR PFM: PFM00001; 7cm.1; 1.  
 DR PRINTS: PR00237; GPCRHHODOPSIN.  
 DR PRINTS: PR00238; GPCRHHODOPSIN.  
 DR PRINTS: PR00579; RHODOPSIN.  
 DR PROSITE: PS00237; G-PROTEIN RECP\_F1\_1; 1.  
 DR PROSITE: PS50262; G-PROTEIN RECP\_F1\_2; 1.  
 DR PROSITE: PS00238; OPSIN; 1.  
 KW Photoceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;  
 KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.  
 FT TRANSMEM 1 36 EXTRACELLULAR.  
 FT TRANSMEM 37 61 1 (POTENTIAL).  
 FT TRANSMEM 62 73 CYTOPLASMIC.  
 FT TRANSMEM 74 98 2 (POTENTIAL).  
 FT TRANSMEM 99 113 EXTRACELLULAR.  
 FT TRANSMEM 114 133 3 (POTENTIAL).  
 FT TRANSMEM 134 152 CYTOPLASMIC.  
 FT TRANSMEM 153 176 4 (POTENTIAL).  
 FT TRANSMEM 177 202 EXTRACELLULAR.  
 FT TRANSMEM 203 230 5 (POTENTIAL).  
 FT TRANSMEM 231 252 CYTOPLASMIC.  
 FT TRANSMEM 253 276 6 (POTENTIAL).  
 FT TRANSMEM 277 284 EXTRACELLULAR.  
 FT TRANSMEM 285 309 7 (POTENTIAL).  
 FT TRANSMEM 310 354 CYTOPLASMIC.  
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
 FT BINDING 15 15 RETINAL CHROMOPHORE.  
 FT LIPID 296 296 PALMITATE (BY SIMILARITY).  
 FT LIPID 322 322 PALMITATE (BY SIMILARITY).  
 FT DISULFID 323 323 BY SIMILARITY.  
 FT MOD RES 110 187 BY SIMILARITY.  
 SQ SEQUENCE 354 AA: 39777 MW: 35C9992101AC919 CRC64;

Query Match 8.4%; Score 128.5; DB 1; Length 354;  
 Best Local Similarity 23.4%; Pred. No. 0.027; Indels 55; Gaps 16;  
 Matches 74; Conservative 57; Mismatches 130;

QY 9 YELLAVIOF---LGIFTNGIIVVNGIDILKHKM-APDLLSCLAVSRIFLOLFTFY 64  
 DB 37 YSILCAVMPILLILGPIINFLVTV-----IQHKLRPLIILLNLAVANFVLCGET 92  
 QY 65 VNYIVYVIFIEFIEFMCANAILLFI---NEELM---LATWLGFFYCAKVASVR---H 112  
 DB 93 VNYSSMNGYFILGATGCVGEPFATLGGELGMSLVLAERYVAVVCKPMSNFRSENH 152

FT	TRANSMEM	114	133	3	(POTENTIAL).
FT	DOMAIN	134	152	4	CYTOPLASMIC.
FT	TRANSMEM	153	176	4	(POTENTIAL).
FT	DOMAIN	177	202	5	EXTRACELLULAR.
FT	TRANSMEM	203	230	5	(POTENTIAL).
FT	DOMAIN	231	252	6	CYTOPLASMIC.
FT	TRANSMEM	253	276	6	(POTENTIAL).
FT	TRANSMEM	277	284	7	EXTRACELLULAR.
FT	DOMAIN	285	309	7	(POTENTIAL).
FT	TRANSMEM	310	348	8	CYTOPLASMIC.
FT	DOMAIN	1	1	9	ACETYLATION. (BY SIMILARITY).
FT	MOD.RES	1	2	10	N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT	CARBOHYD	2	15	11	N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT	CARBOHYD	15	15	12	(BY SIMILARITY).
FT	DISULFID	110	187	13	BY SIMILARITY.
FT	BLINDING	296	296	14	RETINAL CHROMOPHORE.
FT	LIPID	322	322	15	PALMITATE (BY SIMILARITY).
FT	LIPID	323	323	16	PALMITATE (BY SIMILARITY).
FT	LIPID	343	343	17	PHOSPHORYLATION (BY RK).
FT	MOD.RES	343	343	18	(BY SIMILARITY).
FT	SEQUENCE	348 AA:	39023 MW:	19	F83010DC40C450D5 CRC64;

[illegible]

```

CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXY-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF055456; AAC12940.1; -.
CC INTERPRO: IPR000276; -.
CC INTERPRO: IPR000732; -.
CC INTERPRO: IPR001760; -.
CC PFAM: PF00001; 7tm1; 1.
CC PRINTS: PRO0237; GPCRHHODPSN.
CC PRINTS: PRO0238; OPSIN.
CC PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
CC PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
CC PROSITE: PS00238; OPSIN; 1.
CC Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
CC phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
CC Acetylation.
KW DOMAIN 1 36 EXTRACELLULAR.
KW TRANSSEM 37 61 1 (POTENTIAL).
KW DOMAIN 62 73 CYTOPLASMIC.
KW TRANSSEM 74 98 2 (POTENTIAL).
KW DOMAIN 99 113 EXTRACELLULAR.
KW TRANSSEM 114 133 3 (POTENTIAL).
KW DOMAIN 134 152 CYTOPLASMIC.
KW TRANSSEM 153 176 4 (POTENTIAL).
KW DOMAIN 177 202 5 (POTENTIAL).
KW TRANSSEM 203 230 6 (POTENTIAL).
KW DOMAIN 231 252 7 (POTENTIAL).
KW TRANSSEM 253 276 8 (POTENTIAL).
KW DOMAIN 277 284 EXTRACELLULAR.
KW TRANSSEM 285 309 7 (POTENTIAL).
KW DOMAIN 310 348 1 (POTENTIAL).
KW MOD_RES 1 2 ACETYLATION (BY SIMILARITY).
KW CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
KW DISULFID 296 296 RETINAL CHROMOPHORE.
KW BINDING 296 296 BY SIMILARITY.
KW LIPID 322 322 PALMITATE (BY SIMILARITY).
KW LIPID 322 322 PALMITATE (BY SIMILARITY).
KW MOD_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).
KW SEQUENCE 348 AA; 39093 MW; 32605DE5ED94723A CRC64;

```

```

DB 200 NESFVIMFVHFTIPLVIFFCYQGVLETKBAADQOES-ATTOKAKEYTRWVIWY 258
OY 231 LSFLLYFSCHMIKVFSLSKFHRRIFPL----FPLVIGIPSGHSLILIGNKLMO 286
DB 259 VAFILCVAPYASVAFYFTHGSDPGIFMTIFSPFAKSSSIY---NPVIYIMNKKQF-- 313
OY 287 NAKFLLHNSCC 298
DB 314 --RNCMLTTLCC 323
RESULT 10
02CL HUMAN STANDARD; PRT; 312 AA.
AC 095371;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OLFACTORY RECEPTOR 2CL.
GN OR2CL.
OS Homo sapiens (Human)
OC Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Centola M., Chen X., Sood R., Deng Z., Aksentjevich I., Blake T.,
RA Rieke D., Chen X., Wood G., Zaks N., Richards N., Krizman D.,
RA Mansfield E., Apostolou S., Liu J., Shatran N., Vedula A., Hamon M.,
RA Cerces A., Kahn T., Guncuo D., Callen D.F., Richards R.I.,
RA Moyzis R.K., Doggett N.A., Collins F.S., Liu P.P.,
RA Fischel-Ghodsian N., Kastner D.L.;
RT Construction of an approximately 700 kb transcript map around the
RL familial mediterranean fever locus on human chromosome 16p13.3."
CC Genome Res. 9:1-21(1998)
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF098664; AAC83557.1; -.
CC INTERPRO: IPR000276; -.
CC INTERPRO: IPR000725; -.
CC PFAM: PF00001; 7tm1; 1.
CC PRINTS: PRO0237; GPCRHHODPSN.
CC PRINTS: PRO0245; OLFACTORYR.
CC PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
CC PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Multigene family; Olfaction.
KW DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
KW TRANSSEM 26 49 1 (POTENTIAL).
KW DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
KW TRANSSEM 58 79 2 (POTENTIAL).
KW DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
KW TRANSSEM 101 120 3 (POTENTIAL).
KW DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
KW TRANSSEM 140 158 4 (POTENTIAL).
KW DOMAIN 159 196 EXTRACELLULAR (POTENTIAL).
KW TRANSSEM 197 219 5 (POTENTIAL).
KW DOMAIN 220 236 CYTOPLASMIC (POTENTIAL).
KW TRANSSEM 237 259 6 (POTENTIAL).
KW DOMAIN 260 272 EXTRACELLULAR (POTENTIAL).
KW TRANSSEM 273 292 7 (POTENTIAL).
KW DOMAIN 293 312 CYTOPLASMIC (POTENTIAL).
KW DISULFID 293 312 BY SIMILARITY.
KW CARBOHYD 97 189 N-LINKED (GLCNAC. . .) (POTENTIAL).

```



RP VARIANTS ADPR A-51; I-104; R-106; G-135; S-140; E-188; M-209 & R-211  
RX MEDLINE=93304432; PubMed=8317502;  
RA Macke J.P., Davenport C.M., Jacobson S.G., Hennessey J.C.,  
RA Gonzalez-Fernandez F., Conway B.P., Heckenlively J., Palmer R.,  
RA Maumenee I.H., Slewing P., Goutas P., Good W., Nathans J.,  
RT "Identification of novel rhodopsin mutations responsible for  
RT retinitis pigmentosa: implications for the structure and function of  
RL rhodopsin.";  
RL Am. J. Hum. Genet. 53:80-89(1993).  
RN [15]  
RP VARIANT ADPR SER-15.  
RX MEDLINE=93357759; PubMed=8353500;  
RA Kranich H., Bartkowiak S., Denton M.J., Krey S., Dickinson P.,  
RA Divigneau C., Gal A.;  
RT "Autosomal dominant 'sector' retinitis pigmentosa due to a point  
RT mutation predicting an Asn-15-Ser substitution of rhodopsin.";  
RL Hum. Mol. Genet. 2:813-814(1993).  
RN [16]  
RP VARIANT CSNB4 GLU-292.  
RX MEDLINE=93364423; PubMed=8358437;  
RA Dryja T.P., Berson E.L., Rao V.R., Orian D.D.;  
RT "Heterozygous missense mutation in the rhodopsin gene as a cause of  
RT congenital stationary night blindness.";  
RL Nat. Genet. 4:280-283(1993).  
RN [17]  
RP VARIANTS ADPR.  
RX MEDLINE=94375083; PubMed=8088850;  
RA Vaitlhaachen R., Berson E.L., Dryja T.P.;  
RT "Further screening of the rhodopsin gene in patients with autosomal  
RT dominant retinitis pigmentosa.";  
RL Genomics 21:461-463(1994).  
RN [18]  
RP VARIANT ADPR THR-44.  
RX MEDLINE=94357587; PubMed=8076945;  
RA Reis C., Antich J., Gean E., Garcia-Sandoval B., Ramos C., Ayuso C.,  
RA Carralio M.;  
RT "Identification of a novel rhodopsin mutation (Met-44-Thr) in a  
RT simplex case of retinitis pigmentosa.";  
RL Hum. Genet. 94:283-286(1994).  
RN [19]  
RP VARIANTS ADPR PHE-110; PRO-131 AND VAL-164.  
RX MEDLINE=95072600; PubMed=7981701;  
RA Fuchs S., Kirnlich H., Denton M.J., Zrenner E., Bhattacharya S.S.,  
RA Humphries P., Gal A.;  
RT "Three novel rhodopsin mutations (C110F, L131P, A164V) in patients  
RT with autosomal dominant retinitis pigmentosa.";  
RL Hum. Mol. Genet. 3:1203-1203(1994).  
RN [20]  
RP VARIANT ADPR GLN-171.  
RX MEDLINE=95078852; PubMed=7987326;  
RA Antinolo G., Sanchez B., Borrego S., Rueda T., Chaparro P.,  
RA Cabeza J.C.;  
RT "Identification of a new mutation at codon 171 of rhodopsin gene  
RT causing autosomal dominant retinitis pigmentosa.";  
RL Hum. Mol. Genet. 3:1421-1421(1994).  
RN [21]  
RP VARIANTS ADPR PHE-127; PRO-131; ASN-178; ARG-267 AND ARG-297.  
RX MEDLINE=95078858; PubMed=7987331;  
RA Souled E., Gerber S., Rozet J.M., Bonneau D., Dufler J.L., Ghazi I.,  
RA Philip N., Soudhane G., Coscas G., Munnlich A.;  
RT "Five novel missense mutations of the rhodopsin gene in autosomal  
RT dominant retinitis pigmentosa.";  
RL Hum. Mol. Genet. 3:1433-1434(1994).  
RN [22]  
RP VARIANTS ADPR ARG-40 AND LYS-216.  
RX MEDLINE=94362717; PubMed=8081400;  
RA Al-Maghtieh M., Inglehearn C., Lunt P., Jay M., Bird A.,  
RA Bhattacharya S.;  
RT "Two new rhodopsin transversion mutations (L40R; M216K) in families  
RT with autosomal dominant retinitis pigmentosa.";  
RL Hum. Mutat. 3:409-410(1994).  
RN [23]  
RP VARIANT ADPR LEU-345.

RX MEDLINE=94321123; PubMed=80455708;  
RA Rosas D.J., Roman A.V., Weissbrodt P., Moeke J.P., Nathans J.;  
RT "Autosomal dominant retinitis pigmentosa in a large family: a  
RL clinical and molecular genetic study.";   
RN Invest. Ophthalmol. Vis. Sci. 35:3134-3144(1994).  
RP [24]  
RP VARIANT ARRP LYS-150.  
RX MEDLINE=95078913; PubMed=7987385;  
RA Kumeatamrakitcavel G., Maw M., Denton M.J., John S., Srikanum C.R.,  
RA Orth U., Oehlmann R., Gal A.;  
RT "Missense rhodopsin mutation in a family with recessive RP.";   
RL Nat. Genet. 8:10-11(1994).  
RN [25]  
RP VARIANT ADPR ALA-347.  
RX MEDLINE=95359993; PubMed=7633434;

Query Match	8.1%;	Score 123.5;	DB 1;	Length 348;
Best Local Similarity	20.8%;	Pred. No. 0.058;		
Matches 68;	Conservative 63;	Mismatches 125;	Indels 71;	Gaps 16;

QY 1 MLESHLLIYFLAVIQDPLLGIFTNGIIVVNGIDILHKKM-APDLLLSCLAVSRIFLQ 59  
|| :: ||| | : : | :| :| :| :| ||| :| :  
Db 39 MLAAYM---FLIIVGSPINFLTYTV-----QHKKEFLPLNILLNLAADLEMV 87

QY 60 LEIFYNVIVAFIEFIEICMSANCAILFI----NEELW----LATWIGVPCAKVASVR 111  
| | : | : | : || | | : :: |  
Db 88 LGGFTSLYTSLHGVEFGPGCNLEGEFATLGGELIMSLVLAIERVVVCKPKMSNR 147

```

09      112  ---HPL---FIW---LKMRIISKLVPMWIIIGSLIYVSMICVFHSKAGFMVDFYLR--- 157
           | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      148  FGENHAIMGVAFWTWALACAPPLAGW-----SRY---IPEGLOQSC 187

```

DQ 158 --KFSQNATIQEKEDLAIQIESFVAEFSVPLLIIEFAVLIIIFSLSGRHTRQMRTNIVAGS 215  
::: : | : || : :: | : : : : : : :  
Db 188 GIDYITLKPEVN---SFVIYMVFVVHTIPIIIFFCYGGLVFYKEAQAQQES-ATT 243

QY 216 RVPGRGAPISALLSLTSLFELLYFSHCKIKVPLSSLKFKHIRRFIL---FFILVICGYPS 271  
          :::||||: : : : : |||  
Db 244 QKAEEVTBVIIMVIAFLICWVPYASVAEYIFTHQGSNFGPIFMITPAFAKSAIY-- 301

Qy 412 GHSLLILGNPKLQNAKFFLLHSKC 298  
          :|:|:|:|:|:|:|:|:|  
Db 302 -NPVIYIMANKQF---RNCMLTTICC 323

ID	OPSD	DELIDE	STANDARD.	DATE.
				2010 11

AC	06/2/91;
15-DEC-1998	(Rel. 37, Created)
15-DEC-1998	(Rel. 37, Last sequence update)
15-JUL-1999	(Rel. 38, Last annotation update)

Delphinus delphis (Saddleback dolphin) (Black sea dolphin).  
Eumetazoa: Metazoa: Chordata: Cranialata: Vertebrata: Euteleostomi: Rhodopsin.

mammalia; Euteria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
 Delphinus.

Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.

- 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
- 1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES VISION IN DIM LIGHT.

-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS. OPSIN SUBFAMILY.

.....



```

Db 76 ANGMHMPFLFCLPLSG--FIFETIYL-TALFLAASIERFLVANH--WYKTRRLG 130
Oy 122 ---ISKLVPMWMLG---SLVAVSMIC--VPHSKYAGFVWVFLKRFSONATIO---KED 170
Db 131 OAGLVAVACWLIASHKGVAVVIEPSGDISHSQ-----GTNCTCYLERRKD 176
Oy 171 TLAT---QIFSVAFESVPLIFLFAVLLIFSLGR---HTROMRVAGSRVPGRCAP 223
Db 177 QLATLLPRLMMAVAVLFWPLITSYCSRLWMLTGRGSHRROR----- 222
Oy 224 ISALL--SLSEFLIX-----FSH-----C-----MKYFLSLKFIHREIFLF 261
Db 223 VAGLLATLLNPLVCFPGYVSHVGYIGESPPAMRIYVTLTLMSCVDPEVYVF 278

RESULT 14
OPSD_PHOVI STANDARD: PRT: 348 AA.
AC 062754:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Phoca vitulina (Harbor seal).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Plinipedia; Phocidae; Phoca.
RN [1]
RP SEQUENCE FROM N.A.
RA Fastick J.I., Robinson P.R.:
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF05317; AAC12764.1; -
DR INTERPRO: IPR000276; -
DR INTERPRO: IPR001760; -
DR INTERPRO: IPR001760; -
DR PFAM: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHHODOPSIN.
DR PRINTS: PR00238; OPSIN.
DR PRINTS: PR00579; RHODOPSIN.
DR PROSITE: PS00237; G-PROTEIN_RECPT_FL_1; 1.
DR PROSITE: PS00262; G-PROTEIN_RECPT_FL_2; 1.
DR PROSITE: PS00238; OPSIN; 1.
DR PROSITE: PS00238; OPSIN; 1.
DR Phocopsin; Retinal protein; Transmembrane; Glycoprotein; Vision;
DR Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
KW Acetylation.
FT DOMAIN 1 36 EXTRACELLULAR.
FT TRANSMEM 37 61 1 (POTENTIAL).
FT TRANSMEM 62 73 CYTOPLASMIC.
FT TRANSMEM 74 98 2 (POTENTIAL).
FT TRANSMEM 99 113 EXTRACELLULAR.
FT TRANSMEM 114 133 3 (POTENTIAL).
FT TRANSMEM 134 152 CYTOPLASMIC.
FT TRANSMEM 153 176 4 (POTENTIAL).
FT TRANSMEM 177 202 EXTRACELLULAR.

```

```

FT TRANSMEM 203 230 5 (POTENTIAL).
FT DOMAIN 231 252 CYTOPLASMIC.
FT TRANSMEM 253 276 6 (POTENTIAL).
FT TRANSMEM 277 284 EXTRACELLULAR.
FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 348 CYTOPLASMIC.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 110 187 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT BINDING 296 296 RETINAL CHROMOPHORE.
FT LIPID 322 332 PALMITATE (BY SIMILARITY).
FT LIPID 323 323 PALMITATE (BY SIMILARITY).
FT MOD_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).
SQ SEQUENCE 348 AA: 38973 MW: 85189CDBCEEFBE8 CRC64;

Query Match 8.1%; Score 122.5; DB 1; Length 348;
Best Local Similarity 19.9%; Pred. No. 0.067;
Matches 65; Conservative 66; Mismatches 125; Indels 71; Gaps 16;

Oy 1 MESHLLIYELLAVIOFLGIPTNGIIVVNGIDIKRKM-APDLLSCIAVRIFLQ 59
Db 39 MAAVW---FLIVIGFPINFLIVTV-----QHKRLTPNTLLNLAVADLEW 87
Oy 60 LRTFYVNVIVIFIEFIMCSANCAILLTI---NELEW---LATWLGVEYCAKVASVR 111
Db 88 FCGFTTLYTSLHGYFVGPGPCNLEGFATLGELIALSVLAIERVYVVCCKPMSNR 147
Oy 112 ----HPL-----PIW---LKMRSKLVPMWMLGSLIYVSMICVPSKYGVWVFLR--- 157
Db 148 FEGENAIINGVGTWWMALACAPLVGM-----SR---IPGMOQSC 187
Oy 158 --KFSQNNATTOKEPTLAIQIFSVAFESVPLIFLFAVLLIFSLGRHRRONTVAGS 215
Db 188 GIDVYTLKPEVNE---SFVIYMFVHFPIPIVIFECQGLTFVKEAQAQOES-ATY 243
Oy 216 RYVGRGAPISALSTLSLFLIYFSHCKMKFLSSIKFIHRRIFL---PIYVIGIYPS 271
Db 244 QKAEKVEYRWYIYVIALIICWVPYASVAFITFHQSGNSGPIFTLPAFAKASTY-- 301
Oy 272 GHSLLILGNPKLNKAMKFLHSKCC 298
Db 302 -NPVIYIMMKQF---RTCMITTLCC 323

RESULT 15
OPSD_PIG STANDARD: PRT: 348 AA.
AC 018766;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO OR RHOL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97475534; PubMed-9335046;
RA Petters R.M., Alexander C.A., Wells K.D., Collins E.B., Sommer J.R.,
RA Blanton M.R., Rojas G., Hao Y., Flowers W.L., Banin E.,
RA Cideciyan A.V., Jacobson S.G., Wong F.:
RT "Genetically engineered large animal model for studying cone
RT photoreceptor survival and degeneration in retinitis pigmentosa.";
RL Nat. Biotechnol. 15:965-970(1997).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.

```

```

Oy      1 MLESHLIYELFLAVIOFLGIFTNGIIVVANGIDILKIRKA-APDLDLLSLAVSRIFLO 59
         ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      39 MLAAYM--FMLIYGLPPINFLLYTV-----QHRKLRTPPLVNTLLMLAAVDLFM 87
Oy      60 LEIFYVNIYIFIEPIIMCSANCAILLET-----NELELM-----LATMGVPEYCAKVASVR 111
         ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      88 FGFGFTLLYLSHOGFVEFGPTGCTEGFPATLGGEIATLSLVLAIERYVAVCKPMSENR 147
Oy      112 ---HPL---FTW---LKRIKISLYPKMILGSLLYVSMICVFSHKYAGFMVPELR-- 157
         ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      148 FGEHNAIMGIALTYVMALACAPPLGW-----SRY-----TPBGLQSC 187
Oy      158 --KFPSONATIQCKDITAIQIFSFVAEFSPVLLIFLAVLLIFSLGHRTRQMRWVAGS 215
         ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

Search completed: March 15, 2001, 13:20:35  
Job time: 1355 sec



Gencore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 12:52:19 ; Search time 66.3 Seconds  
(without alignments)  
306.219 Million cell updates/sec

Title: US-09-510-332-1

Perfect score: 1521  
Sequence: 1 MESHIIITFLAVIQFLG.....GNPKLNKAKFLHSSKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR66:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131	8.6	333	2 S52960	NADH dehydrogenase
2	130.5	8.6	348	1 JC4267	opsin - rabbit
3	129.5	8.5	348	1 OOSH	rhodopsin - sheep
4	128.5	8.4	348	1 S23398	rhodopsin - Chines
5	125.5	8.3	348	1 S51677	rhodopsin - rat
6	125	8.2	382	2 T26188	hypothetical prote
7	124.5	8.2	445	2 E22845	hypothetical prote
8	124	8.2	500	2 D58931	NADH dehydrogenase
9	123.5	8.1	348	1 OOHU	rhodopsin - human
10	122.5	8.1	335	2 161746	phomomone receptor
11	122.5	8.1	346	2 JC5715	G protein-coupled
12	122.5	8.1	347	2 T27253	hypothetical prote
13	121.5	8.0	348	1 A23665	opsin mouse
14	121	8.0	355	2 151319	Rh2 opsin - green
15	121	8.0	512	1 A70201	virulence factor m
16	120.5	7.9	346	2 JC5716	G protein-coupled
17	120	7.9	349	2 A45229	opsin, green-sensi
18	119.5	7.9	353	1 JN0120	glucose transporte
19	118.5	7.8	524	2 A31318	glucose receptor su
20	118	7.8	420	2 151666	glucose transport
21	117.5	7.7	522	2 A31556	glucose transport
22	117.5	7.7	522	2 S68362	NADH dehydrogenase
23	116.5	7.7	286	1 OXFE2M	angiotensin II rec
24	116.5	7.7	359	2 S15403	cytochrome C-type
25	116.5	7.7	620	2 A58932	rhodopsin - bovine
26	115.5	7.6	348	1 OOB0	rhodopsin - green
27	115.5	7.6	352	2 150081	opsin, green-sensi
28	115.5	7.6	355	2 A42347	opsin, rod - dog
29	115	7.6	312	1 S32696	

30	115	7.6	538	2 EB1435	probable iron-upta
31	114.5	7.5	354	1 151200	rhodopsin - Africa
32	114	7.5	262	2 S59078	conserved hypothet
33	114	7.5	770	2 G72589	hypothetical prote
34	113.5	7.5	345	1 NMBE17	38.1k membrane pro
35	113.5	7.5	359	2 JC1104	angiotensin II rec
36	112	7.4	269	2 D70454	hypothetical prote
37	112	7.4	320	2 S20573	olfactory receptor
38	112	7.4	523	2 S06920	glucose transport
39	111.5	7.3	274	2 S42172	NADH dehydrogenase
40	111.5	7.3	354	2 S27231	rhodopsin - northe
41	111.5	7.3	355	1 A46191	iodopsin homolog
42	111.5	7.3	359	2 JC1194	angiotensin II rec
43	111.5	7.3	359	2 A42656	angiotensin II rec
44	111.5	7.3	468	2 A72619	probable NADH dehy
45	111	7.3	274	2 S42168	NADH dehydrogenase

## ALIGNMENTS

RESULT	1	Query Match	8.6%	Score 131;	DB 2;	Length 333;
S52960	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - honeybee mitochondrion	Best Local Similarity 19.9%; Pred. No. 0.001;				
C:Species:	mitochondrion Apis mellifera (honeybee)	Matches 67; Conservative 63; Mismatches 130; Indels 76; Gaps 13;				
A:Variety:	Ligustica					
C:Date:	28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-2000					
C:Accession:	S52960					
R:Crossier, R.H.; Crozier, Y.C.	Genetics 133, 97-117, 1993					
A:Title:	The mitochondrial genome of the honeybee Apis mellifera: complete sequence a					
A:Reference number:	S52960; WUID:93114603					
A:Accession:	S52960					
A>Status:	preliminary					
A:Molecule type:	DNA					
A:Residues:	1-333 <CHO>					
A:Cross-references:	EMBL:L06178					
C:Genetics:						
A:Genome:	mitochondrion					
A:Genetic code:	SGC4					
C:Superfamily:	NADH dehydrogenase (ubiquinone) chain 2					
C:Keywords:	membrane-associated complex; mitochondrion; NAD; oxidoreductase					
Query Match	8.6%; Score 131; DB 2; Length 333;					
Best Local Similarity	19.9%; Pred. No. 0.001;					
Matches	67; Conservative 63; Mismatches 130; Indels 76; Gaps 13;					
QY	5 HLITVFLAV-----IQFLGIFTNGIIVVNGIDLKRRKAPDLLSCLAV 53					
DB	9 HMEIYFLITFVLMNNSNFIOMNMEF--GTITISLIN-ISTNKPRLIYYSVI 65					
QY	54 SRFLQLEFFVYVNVIFIEFIMCSANCAILFEIN-----ELELV-----LATWLGVEY 103					
DB	66 SSTFL-FEMITVYLSISTKTDTNFEMQMFLKIGPPEHFMIYSYEMNNKQIFL 124					
QY	104 CAVVASVRHPLFIWLK-RISKLVPMILGSLIYSMLCVFHSKAGMVPYFLKRFSSQ 162					
DB	125 MSTLILF-IPYVMVSMKINSMTLYFLITNSLYIS-----FYANKFTTL 168					
QY	163 NATIQEDPLAIOFSEVAD-----FSVPLLFLEFAVLILFSGRHTRQMRNTVAGSRV 217					
DB	169 KKLACSTTFNSFYFLIELNKNMFIAMITIXSFVFLILSFLKNFNQNNFMYNK- 227					
QY	218 PGRGAPISALSLISFLIYFSHCMTKVELS-----SLKFHTRRFLFPIYVIGIY 269					
DB	228 -----YQMTFLLTLMFNYSMPFLSFYIKNLLTFMVSVAVMILFLMISSM 277					
QY	270 PSGLSLILILGNPKLN-----AKKFLHS 295					
DB	278 IMIWMYIILKRVFLKMKNFYKNNPIDDKNKMYHS 313					

## RESULT 2

JC4267

opsin - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: JC4267; PC4072

Gene 162, 331-332, 1995

A:Title: The deduced amino-acid sequence of opsin from rabbit rod photoreceptors.

A:Reference number: JC4267; MUID:96032368

A:Accession: JC4267

A:Molecule type: mRNA

A:Residues: 1-348 &lt;SM2&gt;

A:Cross-references: GB:U12688; NID:9710429; PIDN:AAA91640.1; PID:9710430

A:Accession: PC4072

A:Molecule type: Protein

A:Residues: 1-39;318-348 &lt;SM2&gt;

A:Experimental source: rod photoreceptor cells

C:Comment: This protein is a component of rhodopsin together with the retinoid chromophore

C:Superfamily: vertebrate rhodopsin

F:37-61/Domain: transmembrane #status predicted &lt;TM1&gt;

F:75-99/Domain: transmembrane #status predicted &lt;TM2&gt;

F:109-133/Domain: transmembrane #status predicted &lt;TM3&gt;

F:152-176/Domain: transmembrane #status predicted &lt;TM4&gt;

F:203-227/Domain: transmembrane #status predicted &lt;TM5&gt;

F:253-277/Domain: transmembrane #status predicted &lt;TM6&gt;

F:286-310/Domain: transmembrane #status predicted &lt;TM7&gt;

F:215/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:116/Binding site: retinal (Lys) (covalent) #status predicted

F:322,333/Binding site: palmitate (Cys) (covalent) #status predicted

F:336,340,342/Binding site: phosphate (Thr) (covalent) (by rhodopsin kinase) #status pre

Query Match 8.6%; Score 130.5; DB 1; Length 348;

Best Local Similarity 21.8%; Pred. No. 0.0012;

Matches 70; Conservative 64; Mismatches 128; Indels 59; Gaps 16;

Query 1 MESHLLIYFLAVIOLGIFNGIIVVNGIDILKRRK-APDLLLSCLAVSRIFLQ 59

Db 39 MAAVM---FLIIVGFPINFLIIVTV-----QHKLRPLNLTLLNLAADLFMV 87

Query 60 LFIFYVNIIVFFIIFMCSANCAILFI---NELEIM---LATMIGVYCAKVASVR 111

Db 88 LGFTTTLTYSLHGYFVGPGCNLEGFATLGEIIMSLVLAIRYVYVCKPKMSNFR 147

Query 112 ----HPL----FTLKKRISKLVPMILGSLTY--SMICVFSKXAGFVNPYFLRFFS 161

Db 148 FGENHAIINGVAFWMI-MALACAP-PLVGSRYIPGMCSCGIDY-----YTLKPEVN 199

Query 162 ONATIQKEDTLAIQIFSVFAESVPLIFLFAVLLIFSGRTRQMRNTVAGSRVPGAG 221

Db 200 NESFV-----IYMFVVFHTIPLIIFFCGQLVFYKEAAAOQES-ATTQAEKE 249

Query 222 APISALLSIISFLIYFSHCMIKIVFLSLKFNIRIFL---FFLIIVIGIYPSGSLIL 277

Db 250 VTRNVIIMVIAFLICWPAVSVAFYFTHQSGNFCPIEMTIPAFKSSSY---NPVIY 306

Query 278 ILGNPKLQNAKKFLLSKCC 298

Db 307 IMMKNKF---RNCMLTTTCC 323

## RESULT 3

rhodopsin - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 18-Aug-1982 #sequence\_revision 30-Sep-1990 #text\_change 07-May-1999

C:Accession: A30407; A90319; A93264; A03155

R:Pappin, D.J.C.; Elipoulos, E.; Brett, M.; Findlay, J.B.C.

Int. J. Biol. Macromol. 6, 73-76, 1984

A:Title: A structural model for ovine rhodopsin.

A:Reference number: A91755

A:Accession: A30407

A:Molecule type: Protein

A:Residues: 1-348 &lt;PAP&gt;

A:Note: no explanation is given for the differences in the sequence as seen in this p

R:Brett, M.; Findlay, J.B.C.

BIOchem. J. 211, 661-670, 1983

A:Title: Isolation and characterization of the CNBR peptides from the proteolytically

A:Reference number: A90319; MUID:83282605

A:Accession: A90319

A:Molecule type: Protein

A:Residues: 1:40-44;45-86;87-111;144-155;156-163;164-183;184-207;208-241 &lt;BRE&gt;

R:Findlay, J.B.C.; Brett, M.; Pappin, D.J.C.

Nature 293, 314-316, 1981

A:Title: Primary structure of C-terminal functional sites in ovine rhodopsin.

A:Reference number: A93264; MUID:82013638

A:Accession: A93264

A:Molecule type: Protein

A:Residues: 240-348 &lt;FIN&gt;

R:Pappin, D.J.C.; Findlay, J.B.C.

BIOchem. J. 217, 605-613, 1984

A:Title: Sequence variability in the retinal-attachment domain of mammalian rhodopsin

A:Reference number: A90324; MUID:84178280

A:Contents: annotation; retinal binding site

R:Thompson, P.; Findlay, J.B.C.

BIOchem. J. 220, 773-780, 1984

A:Title: Phosphorylation of ovine rhodopsin: Identification of the phosphorylated site

A:Reference number: A4548; MUID:84279984

A:Contents: annotation; phosphorylation sites

C:Superfamily: vertebrate rhodopsin

C:Keywords: acetylated amino end; chromoprotein; eye; G protein-coupled receptor; gly

sion

F:37-61/Domain: transmembrane #status predicted &lt;TM1&gt;

F:114-133/Domain: transmembrane #status predicted &lt;TM2&gt;

F:153-175/Domain: transmembrane #status predicted &lt;TM3&gt;

F:203-230/Domain: transmembrane #status predicted &lt;TM4&gt;

F:253-276/Domain: transmembrane #status predicted &lt;TM5&gt;

F:285-309/Domain: transmembrane #status predicted &lt;TM6&gt;

F:215/Binding site: acetylated amino end (Met) #status experimental

F:296/Binding site: retinal (Lys) (covalent) #status predicted

F:322,333/Binding site: palmitate (Cys) (covalent) #status experimental

F:334,338,343/Binding site: phosphate (Ser) (covalent) (by rhodopsin kinase) #status

F:335,336/Binding site: phosphate (Thr) (covalent) (by rhodopsin kinase) #status expe

rience

Query Match 8.5%; Score 129.5; DB 1; Length 348;

Best Local Similarity 21.1%; Pred. No. 0.0014;

Matches 69; Conservative 62; Mismatches 125; Indels 71; Gaps 16;

Query 1 MESHLLIYFLAVIOLGIFNGIIVVNGIDILKRRK-APDLLLSCLAVSRIFLQ 59

Db 39 MAAVM---FLIIVGFPINFLIIVTV-----QHKLRPLNLTLLNLAADLFMV 87

Query 60 LFIFYVNIIVFFIIFMCSANCAILFI---NELEIM---LATMIGVYCAKVASVR 111

Db 88 LGFTTTLTYSLHGYFVGPGCNLEGFATLGEIIMSLVLAIRYVYVCKPKMSNFR 147

Query 112 ----HPL----FTLKKRISKLVPMILGSLTY--SMICVFSKXAGFVNPYFLRFFS 161

Db 148 FGENHAIINGVAFWMI-MALACAP-PLVGSRYIPGMCSCGIDY-----YTLKPEVN 199

Query 162 ONATIQKEDTLAIQIFSVFAESVPLIFLFAVLLIFSGRTRQMRNTVAGSRVPGAG 221

Db 200 NESFV-----IYMFVVFHTIPLIIFFCGQLVFYKEAAAOQES-ATTQAEKE 249

Query 222 APISALLSIISFLIYFSHCMIKIVFLSLKFNIRIFL---FFLIIVIGIYPSGSLIL 277

Db 250 VTRNVIIMVIAFLICWPAVSVAFYFTHQSGNFCPIEMTIPAFKSSSY---NPVIY 306

Query 278 ILGNPKLQNAKKFLLSKCC 298

Db 307 IMMKNKF---RNCMLTTTCC 323

Query 158 --KFSFNATIQKEDTLAIQIFSVFAESVPLIFLFAVLLIFSGRTRQMRNTVAGS 215

Db 188 GALYFTLKPETINNE---SEVYIMFVHFISPLIYIFCYGQLVFTVEAAAOQES-ATT 243

Query 216 RVPRGAPISALLSIISFLIYFSHCMIKIVFLSLKFNIRIFL---FFLIIVIGIYPS 271

Db 244 OKAEKVTNVIIMVIAFLICWPAVSVAFYFTHQSGDPGFEMTIPAFKSSSY-- 301

QY 272 GHSLLILGNPKLKNKKFLLHSKCC 298  
Db 302 -NPVLYIMNKKF---RNCMLTTLCC 323

## RESULT 4

S23398

rhodopsin - Chinese hamster

C:Species: Cricetus griseus (Chinese hamster)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S23398

R:Gale, J.M.; Tobey, R.A.; D'Anna, J.A.

J. Mol. Biol. 224, 343-358, 1992

A:Title: Localization and DNA sequence of a replication origin in the rhodopsin gene loc

A:Reference number: S23398; MID:92219256

A:Accession: S23398

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-348 &lt;GAL&gt;

C:Cross-references: EMBL:X61084; NID:949478; PIDN:CAA43398.1; PID:949479

C:Genetics:

A:Introns: 121/1; 177/2; 232/3; 312/3

C:Superfamily: vertebrate rhodopsin

C:Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; F

F:37-61/Domain: transmembrane #status predicted &lt;TM1&gt;

F:74-96/Domain: transmembrane #status predicted &lt;TM2&gt;

F:114-133/Domain: transmembrane #status predicted &lt;TM3&gt;

F:153-175/Domain: transmembrane #status predicted &lt;TM4&gt;

F:203-220/Domain: transmembrane #status predicted &lt;TM5&gt;

F:253-276/Domain: transmembrane #status predicted &lt;TM6&gt;

F:285-309/Domain: transmembrane #status predicted &lt;TM7&gt;

F:296/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 8.4%; Score 128.5; DB 1; Length 348;

Best Local Similarity 21.8%; Pred. No. 0.0017;

Matches 70; Conservative 63; Mismatches 129; Indels 59; Gaps 16;

QY 1 MESHLLIYFLAVIOFLGIFTNGIIVVNGIDIKRKM-APLDLISCLAVSRIFLQ 59

Db 39 MAAVM---FLIIVGFPINFLTLVTV-----QHKRLTPNLTLLNLAVADLFMV 87

QY 60 LFIYVNVIVIFIEFIMCSANCALLFL---NELIEM---LATWLGVEYCAKVASVR 111

Db 88 FGGFTTTLTSLHGVEFGPTGCLNLEGFPATLGEIGLSLVLAIERVVVCKPMNSFR 147

QY 112 ----HPL----FIVLKMRIKSLVPMILGSLIV--SMICVFHSKYAGFMVYFLRKEFS 161

Db 148 FGENHAIMGVAFWVMAACAAP-PLVGMSTRYIPEGMCSCGVDY-----YTLKPEVN 199

QY 162 QNATIQEDTLAIQIFSEVAEFSVPLILFVAVLLIFSLGKHTQMRNTVAGSRVPGK 221

Db 200 NESFV-----IYMFVVFHTIPLIVIFFCYGOLVFVKBAADQOQS-ATTQAEKE 249

QY 222 APTSALLISFLILYFESHQMIKVFSLKFKHRRIFL---FELVIGIYPSGSHLTL 277

Db 250 VTRMVLMAVFFLICWFPAGVAFYFTHOGSNFGPIFMTLPAFAFKSSSY--NPVLY 306

QY 278 ILGNPKLKNKKFLLHSKCC 298

Db 307 IMNKKF---RNCMLTTLCC 323

## RESULT 5

S51677

rhodopsin - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S51677

R:Huber, A.; Baker, B.B.; Sander, P.; Gerdon, G.; Paulsen, R.; Williams, T.P.

Submitted to the EMBL Data Library, December 1994

A:Description: Light-history effects: Levels of rhodopsin, opsin and opsin mRNA in albin

A:Reference number: S51677

A:Accession: S51677  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-348 <HOB>

A:Cross-references: EMBL:Z46957; NID:9603874; PIDN:CAA87081.1; PID:9603875

C:Superfamily: vertebrate rhodopsin

C:Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprote

F:296/Binding site: retinal (Lys) (covalent) #status predicted

F:322/323/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 8.3%; Score 125.5; DB 1; Length 348;

Best Local Similarity 20.2%; Pred. No. 0.003;

Matches 66; Conservative 65; Mismatches 125; Indels 71; Gaps 16;

QY 1 MESHLLIYFLAVIOFLGIFTNGIIVVNGIDIKRKM-APLDLISCLAVSRIFLQ 59

Db 39 MAAVM---FLIIVGFPINFLTLVTV-----QHKRLTPNLTLLNLAVADLFMV 87

QY 60 LFIYVNVIVIFIEFIMCSANCALLFL---NELIEM---LATWLGVEYCAKVASVR 111

Db 88 FGGFTTTLTSLHGVEFGPTGCLNLEGFPATLGEIGLSLVLAIERVVVCKPMNSFR 147

QY 112 ----HPL----FIVLKMRIKSLVPMILGSLIVSMICVFHSKYAGFMVYFLR--- 157

Db 148 FGENHAIMGVAFWVMAACAAPPLVGM-----SRY-----IPEGMCSC 187

QY 158 --KFSQNTIQEDTLAIQIFSEVAEFSVPLILFVAVLLIFSLGKHTQMRNTVAGS 215

Db 188 GIDYTLKPEVNE---SEVYMFVVFHTIPLIVIFFCYGOLVFVKBAADQOQS-ATT 243

QY 216 RVPGRGAPISALLISFLILYFESHQMIKVFSLKFKHRRIFL---FELVIGIYPS 271

Db 244 QKAKEVTRVVIIMVIFFLICWLPVAVAMVIFTHOGSNFGPIFMTLPAFAFKTASY-- 301

QY 272 GHSLLILGNPKLKNKKFLLHSKCC 298

Db 302 -NPVLYIMNKKF---RNCMLTTLCC 323

RESULT 6

T26188

hypothetical protein W05B5.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T26188

R:McLay, K.

Submitted to the EMBL Data Library, November 1996

A:Reference number: T20167

A:Accession: T26188

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-382 &lt;WIL&gt;

A:Cross-references: EMBL:Z82071; PIDN:CA804918.1; GSPDB:GN00019; CESP:W05B5.2

A:Experimental source: clone W05B5

C:Genetics:

A:Gene: CESP:W05B5.2

A:Map position: 1

A:Introns: 48/2; 97/3; 146/3; 185/1; 216/3; 268/2; 305/2

C:Superfamily: neurokinin 1 receptor

Query Match 8.2%; Score 125; DB 2; Length 382;

Best Local Similarity 21.5%; Pred. No. 0.0035;

Matches 60; Conservative 54; Mismatches 105; Indels 60; Gaps 11;

QY 8 IYFLAVIOFLGIFTNGIIVV-----NGIDIKRKMAPLDLISL 50

Db 50 VITLASM--VIGVGNLTLLVVVATNKSVMGRMGTRGNKRVVCSMPKRNALMLVLNM 107

QY 51 LAVSRIFQLFTFYVNVIVIFIEFIMCSANCALLFLINELEMLATVGLGYF-CAKVAS 109

Db 108 LVIADLLIFLPLVNVVDVTKTFMFSVAVFCKSVFNNTSVVSVIMSLVFTICERRA 167



```

OY      216 RVEGRGAPISALLSIFSLFYVSHCMKVFSLSKFHRRFLL-----FFILVIVITPS 271
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      244 OKAEKEVTRVYIIVLVIAFLICWVPYASVAFYIFTHOCSNGPIFMPIPAFFAKSAAY-- 301
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      272 GHSLLILGNPKLKNQAKFLLHSKCC 298
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      302 -NPVIYIMANKQF-----RNCMLTFTIC 323
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
161746
pheromone receptor VN4 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 29-Sep-1999
C:Accession: 161746
R:Duclac, G.; Axel, R.
Cell 83, 195-206, 1995
A:Title: A novel family of genes encoding putative pheromone receptors in mammals.
A:Reference number: A57223; MUID:96028094
A:Accession: 161746
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-335 <RES>
A:Cross-references: EMBL:U03696; NID:g1055249; PIDN:AAC52285.1; PID:g1052520
C:Superfamily: pheromone receptor VN3t

Query Match      8.1%; Score 122.5; DB 2; Length 335;
Best Local Similarity 20.8%; Pred. No. 0.0051;
Matches 63; Conservative 64; Mismatches 135; Indels 41; Gaps 11;

OY      19 LGIFFTNGIIVVNGIDILKRRMAPDLLLSCLAVSRIFLQLFFIVVYVYFFIEFTMC 78
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      48 VGIANSILFFGHLCLMLGCKNPRIHLYIASLSLTQLML-LITMGLAADFFISQGIWD 106
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      79 SANCALILFINELE-----LMLATWIGVFYCAVASVRHPLEFTLMKRISKLVPMMILGSL 134
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      107 STSCOSLILYLRHSRGFTLSAACLNVFMITLSKRKSL-----TKFKHNSPHHISGAF 161
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      135 LVYSMICVFHSKXAGFVVPYFLRKFFSQNAT-----IQKEDTLAIQIFSAFESVPLL 188
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      162 L---LLCVLYMCCSHLI---LSIIATPNLTSDFNMYVYKSCSFLPWCYSRYSMEFTTIA 215
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      189 I---FLPAVL-----LLITSIGHTRQMRNT-----VAGSRVGRGAPISALLSIFLIL 236
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      216 VREAFPIGMLALSGYLVAFLMRHRQAOQLHSTGTSKSSPEQRAETETILLMSFFVL 275
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      237 YFHCMIKVPYLSLKFIHRRIFLFFILVIGIYPSCHSLILIGNPKLKNQAKFLLHSK 296
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      276 YLLENV--VFYSMKRRKDSFTYCYQILIVSHATVATSSPFIFETEKRMK-----ILRSV 328
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      297 CCQ 299
           : :
DB      329 CAR 331

RESULT 11
JC5715
G protein-coupled receptor 41 - human
C:Species: Homo sapiens (man)
C:Date: 03-Dec-1997 #sequence_revision 03-Dec-1997 #text_change 29-Sep-1999
C:Accession: JC5715
R:Samadargu, M.; George, S.R.; Nguyen, T.; Xu, S.; Kolakowski Jr., L.F.; O'Dowd, B.F.
Biochem. Biophys. Res. Commun. 239, 543-547, 1997
A:Title: A cluster of four novel human G protein-coupled receptor genes occurring in
A:Reference number: JC5714; MUID:96008875
A:Accession: JC5715
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-346 <SNM>
A:Cross-references: GB:AF024688; NID:g2612947; PIDN:AAB86711.1; PID:g2612948
C:Superfamily: G protein-coupled receptor 43

```

379 LIPEDDARQFALT LHPCTREFFTSEVLLIADNIIERFIWIFPIEGVYFASFPLFO 438

—  
• •  
—  
• •  
• •  
• •  
—  
• •  
• •  
—  
—

Db 307 IM - - - - LNKQFRNCMLTTLCC 323







```

484 .....CAGCACAGAGAGCTGCGCAGCGCTCTCAACTACATCTCTGTC 525
50 CysleuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValas 66
|||||.....:|||||.....:|||||
526 AACCTAGCGGTGGCTGACCTCTTCATGCTCTCAAGTGGTCTTACCCAGCAG 575
66 nValIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysA 83
|||||.....:|||||.....:|||||
576 CCTCTACACCTCTCTGATGATGATCTCTCTCTGCGGCCACAGATGCA 625
83 IalleLeuLeuPheIle.....AsnGluLeuGluLeuTyr... 94
626 ATTTGGAGGGCTTTTGGCCACCTGGGGGGAATTCCTGCTGCTGTC 675
95 .....LeuAlaThrTyrLeuGluGlyValPheTyrCysAlaIleValAl 108
|||||.....:|||||.....:|||||
676 TTGCTGCTCTGCGCATCGAGCGGTGAGTGGTGTGTGTGAACCCATGAG 725
108 aserValArg.....HisProLeu.....PheIleT 117
|||||.....:|||||.....:|||||
726 CAACCTCCGCTTCGGGAGAGACCATCCATGCGCGCTTGCCTTACCT 775
117 TP.....LeuLysMetArgIleSerLysLeuValProThrMetIle 130
|||||.....:|||||.....:|||||
776 GGGTCATGGCGCTGGCTCGCCGCGCACCCCTCGCGCGCTG..... 819
131 LeuGlySerLeuLeuTyrValSerMetIleCysValPheHisSerLysTyr 147
|||||.....:|||||.....:|||||
820 .....TCCAGGTA 827
147 ValGluPheMetValProTyrPheLeuArg.....LysP 159
|||||.....:|||||.....:|||||
828 C.....ATCCCGAGGGCTCGAGTCTGCTGTGGAATGACT 865
159 hePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeuAlaIleGln 175
|||||.....:|||||.....:|||||
866 ACTACACGCTCAAGCGGAGTCAACAAGAG.....TCTTTTCTC 906
176 IlePheSerPheValAlaGluPheSerValProLeuLeuIlePheLeuPh 192
|||||.....:|||||.....:|||||
907 ATCTACATGCTGTGCTCACTTCACTCCATCCCATGATTAATCATCTTTT 956
192 eAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArgIleMetA 209
|||||.....:|||||.....:|||||
957 CTGCTATGGGCGCTGCTCTTCAACGCTCAAGAGGCGCTCCACGACGC 1006
209 rGAsnThrValAlaGlySerArgValProGluArgGlyAlaProIleSer 225
|||||.....:|||||.....:|||||
1007 AGGAGTCA...GCCACACACAGAGAGGAGAGAGAGTCAACCGGATG 1053
226 AlaLeuLeuSerIleLeuSerPheLeuIleLeuTyrPheSerHisCysMe 242
|||||.....:|||||.....:|||||
1054 GTCATCATCATGTCATGCTCTTCTATCTGCTGGGCGCTTACGCCAG 1103
242 IleLeuValPheLeuSerSerLeuLysPheHisIleArgThrPheIleP 259
|||||.....:|||||.....:|||||
1104 CGTGCCATTTACATCTTCAACGAGGCTCCAACTTCCGATCCCATCT 1153
259 heLeu.....PhePheIleLeuValIleGlyIleTyrProSer 271
|||||.....:|||||.....:|||||
1154 TCATGACATCCAGCGCTTCTTGGCAAGAGCGCGCATCTAC..... 1197
272 GlyHisSerLeuIleLeuGluIleValAsnProLysLeuGlnAsnAl 288
|||||.....:|||||.....:|||||
1198 ...AACCGTGCATCTATATCATGATGACACAGCAGTTC..... 1233
288 eAlaLysPheLeuLeuLeuHisSerLysCysCysGln 299
|||||.....:|||||.....:|||||
1234 .CGGACATGCTGCTCAACACCATCTGCTGCGCG 1266
seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-033-081B-1
seq_documentation_block:

```

```

Sequence 1, Application US/08033081B
Patent No. 5498521
GENERAL INFORMATION:
APPLICANT: Dryja, Thadeus P.
APPLICANT: Berson, Eliot L.
TITLE OF INVENTION: DIAGNOSIS OF HEREDITARY RETINAL
DEGENERATIVE DISEASES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/033,081B
FILING DATE: March 11, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/825,296
FILING DATE: January 23, 1992
APPLICATION NUMBER: 07/469,215
FILING DATE: January 24, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/069005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3016
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-033-081B-1

alignment_scores:
Quality: 124.50 Length: 328
Ratio: 0.759 Gaps: 16
Percent Similarity: 50.000 Percent Identity: 20.732

alignment_block:
US-09-510-332-1 x US-08-033-081B-1 ..

Align seg 1/1 to: US-08-033-081B-1 from: 1 to: 3016

1 MetLeuGlnSerHisLeuIleIleTyrPheLeuLeuAlaValIleGlnPh 17
|||||.....:|||||.....:|||||
409 ATGCTGCGCGCTTACATG.....TTTCTGATGATGCTGCTGAGCTT 449
17 eLeuLeuGlyIlePheThrAsnGlyTyrIleLeuValValaAsnGlyIleA 34
|||||.....:|||||.....:|||||
450 CCCCATCACTTCTCTACGCTCTACGTCACCGCTC..... 483
34 sPheuIleLysHisArgLysMet...AlaProLeuAspLeuLeuSer 49
|||||.....:|||||.....:|||||
484 .....CAGCACAGAGAGCTGCGCAGCGCTCTCAACTACATCTCTGCTC 525
50 CysLeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValas 66
|||||.....:|||||.....:|||||
526 AACCTAGCGGTGGCTGACCTCTTCAAGTGGTCTTACCCAGCAG 575
66 nValIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysA 83

```

```

576 CCTCACACCTCTGTGCATGATCTCTTCCGCGCCACAGAGATCA 625
83 lalleleuLeuphele.....AangluLeugluLeuTri... 94
626 ATTGGAGGCGCTTCTTGTCCACCCCTGGGGGTGAATTGCCCTGTGTCC 675
95 .....LeuAlaThrTrpleugluValPheTYCysAlaValAl 108
676 TTGGTGCTGCTGTGCATGAGCGGTAGCTGGGTGTAGAGCCATGAG 725
108 aSeValArg.....HisProLeu.....PheIleT 117
726 CAACCTTCGCTTCGGGAGAACCATGCATGCGCTTGCCTTCACCT 775
117 TP.....LeuLysMetArgIleSerLysLeuValProTriPheTle 130
776 GGGTCATGGCGCTGCTGCGCCGACCCCTGCGCGCTG..... 819
131 LeuGlySerLeuLeuTYValSerMetIleCysValPheHisSerLysTY 147
820 .....TCCAGGTA 827
147 rAlaGlyPheMetValProTYrPheLeuArg.....LysP 159
828 C.....ATCCCGAGGCGCTGCATGCTGCTGTGGAATGACT 865
159 hePheSerGlnAsnAlaThrIleGlnLysGlnAspThrLeuAlaIleGln 175
866 ACTACACGCTCAAGCCGAGAGTCACACAGAG.....TCTTTGTC 906
176 lLePheSerPheValAlaGluPheSerValProLeuLeuIlePheLeuPh 192
907 ATCTACATGTTCTGTGCTCCTTACACCATCCCATGATATCATCTTTT 956
192 eAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArgGlnMetA 209
957 CTGCTATGGCGACTGCTCTTCAACCGCAAGAGGCCGCTGCCACAGC 1006
209 rGAsnThrValAlaGlySerArgValProGlyArgGlyAlaProIleSer 225
1007 AGGAGTCA...GCCACACACAGAGAGGACAGAGAGGACCCGGATG 1053
226 AAlaLeuSerIleLeuSerPheLeuIleLeuTYrPheSerHisCysMe 242
1054 GTCATCATCATGATCATGCTTCTCTGATCTGCTGGGCTTACAGCCAG 1103
242 tIleLeuValPheLeuSerSerLeuLysPheHisIleArgArgPheIleP 259
1104 CGTGCAATTTCTACATCTTCAACCCACAGGGCTCCAACTTGGTCCCATCT 1153
259 heLeu.....PhePheIleLeuValIleGlyIleTYrProSer 271
1154 TCATGACCATCCACAGCGTTCTTTCACAGAGCGCCCATCTAC..... 1197
272 GlyHisSerLeuIleLeuIleLeuGlyAsnProLysLeuLysGlnAsnAl 288
1198 ..AACCCTGTCATCTATATCATGATGATGAACAAGAGTTC..... 1233
288 aLysLysPheLeuLeuHisSerLysCysCysGln 299
1234 .CGGAATGCAATGCTCACCACCATCTGCTGCCCG 1266
seq_name: /cgn2_6/ptodata/2/1na/6_COMB.seq:US-08-931-999-4
seq_documentation_block:
: Sequence 4, Application US/08931999
: Patent No. 6043219
: GENERAL INFORMATION:
: APPLICANT: Iandolo, John J.
: APPLICANT: Crupper, Scott S.
: TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
: NUMBER OF SEQUENCES: 4

```

```

CORRESPONDENCE ADDRESS:
ADDRESSER: Hovey, Williams, Timmons & Collins
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: Missouri
COUNTRY: U.S.A.
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,999
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/710,561
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 25043-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 816/474-9050
TELEFAX: 816/474-9057
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6755 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
TOPOLOGY: unknown
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: UT0007
US-08-931-999-4

alignment_scores:
Quality: 124.50 Length: 238
Ratio: 1.029 Gaps: 13
Percent Similarity: 50.840 Percent Identity: 26.471

alignment_block:
US-09-510-332-1 x US-08-931-999-4/rev ..
Align seg 1/1 to reverse of: US-08-931-999-4 from: 1 to: 6755

48 leuSerCysLeuAlaValSerArgIlePheLeuGlnLeuPhe.....11 62
|||||.....:|||||:|||||
908 CTCTCTGGGCTGCGCTTTTGGTGTGCTTTTGGTGGGCTT 859
62 ePheTYrValAsnValIleValIlePhePheIleGluPheIleMetCys 79
|||||.....:|||||:|||||
858 CTGTGTTTGGGGGTGTTTGTGTTTTCGTCCTTTCGCTGTT 809
79 eAlaAsnCysAlaIleLeu.....PheIleAsnGluLeu 93
|||||.....:|||||:|||||
808 TTTGTTGCTGGGGTTCCTGTTCTTCTGTTTCTTTTGGCTTTT 759
94 TrPheAlaThrTrPleugluValPheTYrCysAlaValAlaSerVal 110
|||||.....:|||||:|||||
758 TTCTGTGCTTTTGGGTGTTTTC..... 729
110 lArgHisProLeuPheIleTrPleuLysMetArgIleSerLysLeuValP 127
|||||
728 .....TGG..... 726
127 roTriPheTleLeuGlySerLeuLeuTYrValSerMetIleCysValPhe 143
|||||.....:|||||:|||||

```

```

725 ..TGGGGCGTTTGG...TCCGTTTGTGTTGCTTCGTTTCTTGCGCTTT 681
144 H1SERLYTYRVALAGLPHEMETVALPROTYRPE..... 155
680 CCGGCTTCCTTC.....TTTTTGTTCTTTTGTTCGTTGGTCCCTC 637
156 .....LeuArgLysPhePheSerGlnAsnAlaIat 165
636 CCCCTTCCTCTGTTGTGTTGCTGCTCTCTTTTCTTTTCTTCCTTC 587
165 hrlleGlnLysGlnAspThrIleAlaIleGlnIlePheSerPheValAla 181
586 CTGTG.....GTGGCTTTTGTGCTTTTCTTTTCTTTTCTTCGCG 552
182 GluPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuLeu1 198
551 CCGTTTGGGTTCCCTTTTGTGCTTCCTTCGCTTCGCTTCGTTTCTTC 502
198 ePheSerLeuGlyArgHisThrArgGlnMetArgAsnThrVal...AlaG 214
501 TTTCTGGGTTTGTCTTTTGTGCTTTCTGCTTTCTCTCTTTCTTTT 452
214 LysArgValProGlyArgGlyAlaProIle.....SerAlaLeuLeu 228
451 TCTCGTCCCCCGCTTCCTGCTGCTGCTGCTTTTCTCTTTTGTGT 402
229 SerIleLeuSerPheLeuIleLeuTyrrPheSerHisCysMetIleLysVa 245
401 TCTCTCTCTGTTGCTGCTTTTCTTTTCTTTTCT.....TTCGCT 361
245 lPheLeuSerSerLeuLysPheHisIleArgArgPhe...IlePheLeuP 261
360 CTTCCTG...CGGCTTGCGTTTCGCGCTTTTCTTTTCTTCGCTTTGT 314
261 hPhePheIleLeuVal 265
313 TTTTCTGCTGCTT 300

seq_name: /cgn2_6/ptodata/2/lna/5b_COMB.seq:US-08-820-521-1

seq_documentation_block:
Sequence 1, Application US/08820521
Patent No. 5942416
GENERAL INFORMATION:
APPLICANT: Bergsma, Derk
APPLICANT: Ganesh, Sathe
APPLICANT: Fuelterer, Wendy
APPLICANT: Mao, Joyce
TITLE OF INVENTION: CDNA CLONE HNFY20 THAT ENCODES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,521
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344

```

```

REFERENCE/DOCKET NUMBER: GH50011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1841 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-820-521-1

```

```

alignment_scores:
Quality: 122.50 Length: 296
Ratio: 0.833 Gaps: 18
Percent Similarity: 49.662 Percent Identity: 23.986

```

```

alignment_block:
US-09-510-332-1 x US-08-820-521-1

```

```

Align seg 1/1 to: US-08-820-521-1 from: 1 to: 1841

```

```

10 PheLeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThrAsnGlyI 26
341 TTTCTGGGTACCTTCACCTTCCTGTTGGGGCTCCCTCAAC...CT 387
26 eIleValAlaValAsnGlyIleAspLeuIleLysHisArgLysMetAlaP 43
388 GCTGGCCCTGCTGCTTCCTGCGGCAAGCTGCACGCCGCGGTGCC. 436
43 roleAspLeuLeuLeuSerCysLeuAlaValSerArgIlePheLeuGln 59
437 ..GTGGACGTGCTCTCTCACTCACTGACCGCTCGACCTGCTGCTG 484
60 LeuPheIlePheTyrrValAsnValIleValIle.....PhePn 72
485 CTGTCTCTGCTTCCTCCGATGCTGGAGCAGCCCAATGCATGCATGCC 534
72 eIleGlnPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIle 89
535 CTTGCCCTTCATCTCTGCGCCACTCTCTGA.....TTCATCTTCTTCA 578
89 snGluLeuGluLeuTyrrPheAlaThrIlePheGlyValPheTyrcysAla 105
579 CCACATCTATCTC...ACGCGCTCTCTCTGACGCTGAGCATTTGAA 625
106 LysValAlaSerValArgHisProLeuPheIleTyrrPheLysMetArg.. 121
626 CGCTTCCTGAGTGTGGCCACCCCTG.....TGTACAAAGCCCGGCC 669
122 .....IleSerLysLeuValProTyrMetIleLeuG 132
670 GAGGCTGGGGCAGCAGGTCGTGTGAGTGTGCTGCTGCTGCTGCT 719
132 Ly.....SerLeuLeuTyrrValSerMetIleCys.....ValPhe 143
720 CTGCTACAGTGCAGCGTGTCTAGCTATAGAAATTCACGAGCATCTCC 769
144 H1SERLYTYRVALAGLPHEMETVALPROTYRPHLEUARGLYSPHEPH 160
770 CACAGCCAG..... 778
160 eSerGlnAsnAlaThrIleGln.....LysGlnAspThrLeuAlaI 174
779 ..GCGACCAATGGGACCTGCTACCTGAGCTTCGGAAGCAGCATAGCA 827
174 le.....GlnIlePheSerPheValAlaIlePheSerValPro 186
828 TCTCTCTGCGCGCTGGGGCTGAGATGCTGTGCTCTCTTGTGTGCTCCG 877

```



```

Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

```

```

alignment_scores:
Quality: 120.50      Length: 263
Ratio: 0.867        Gaps: 14
Percent Similarity: 52.852      Percent Identity: 22.433

```

```

alignment_block:
US-09-510-332-1 x US-08-770-379-20 ..

```

```

Align seg 1/1 to: US-08-770-379-20 from: 1 to: 32207

```

```

11 LeuenaValIleGlnPheLeuLeuGlyLePheThrSnglyIleI 27
24227 ATACTCTCTGATT...TTCTCTATAAGCTTCTGGAATGATTCGT 24273
27 evalValValasnGlyLeaPheLeuIleYSHsArgLysMetAla...P 43
24274 CACCTACAT...TTTGCAMGACCGCATCCGGGCGACAGAG 24311
43 roLeuaPheLeuLeuSerCysLeuAlaValSerArgIlePheLeuGln 59
24312 CGATAGATATACCTCTCTGGTATCTGCTTAACCTGCTGCTTAGC 24361
60 LeuPheIlePheYValAsnValIleValIlePhePheIleGlnPheI 76
24362 ATATCTCTA...TTGGCAGAAGGTTGATGTTTGTTCCTCAATAT 24405
76 eMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnGluLeuL 93
24406 CACTCCACAGCGCTTGCGAGACTTGAAATTTT...T 24443

```

```

93 euTriPheAlaThrTriPheLeuGlyValPheTyrcysAlaLysValAlaSer 109
24444 ACTATTATATCTACTATGATATCTACAGTGTGTGCTGCTGCTGCTA 24493
110 ValArgHisProLeuPheLeu...TriPheLysMetArg 121
24494 GTGAGGTACTCTCTGTCGATATTTACGGCTTCCTGCCCAAG... 24538
121 gIleSerLysLeuValProTriPheIleLeuGlySerLeuLeuValS 138
24539 ...AAGCAGTCCCTCGGATGAGTACATCCGCTGCTGCTGCTGCTA 24584
138 emetIle...CysValPheHisSerLysTyrcysAlaGly 149
24585 CATTGCTGCTCTCGGGGATGCCCTGTCACACAGAGAGGCGGTGTCAC 24634
150 PheMetValProTyrcPheLeuArgLysPhePheSerGlnAsnAlaThrI 166
24635 ...CCGTCACAGACAGCCATGCTGTTATGAAACGGCGGAA 24675
166 eGlnLysGluAsp...ThrLeuAlaIleGlnIlePheSerPheValAlaG 182
24676 CATGACTGCAGACTCGGAGCTGATGTCACAAACCGTCTGCTGCTG 24725
182 LysPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuIle 198
24726 GTTTCCTCTTACCCCTG...GCCCTCTTATTTCTGTT 24760
199 PheSerLeuGly...ArgHisThrArg...GlnMetArgas 210
24761 TATCTCTCACTGCTGCTGCTGAGGAGGACAAAGCTGCAAGCAGCG 24810
210 nTriValAlaGlySerArgValProGlyArgGlyAlaProIleSerAla 227
24811 GAAGGTAAAGGG... 24823
227 euLeuSerIleLeuSerPheLeuIleLeuTyrcPheSerHisCysMetIle 243
24824 ...GTGATTGTTGCTGCTGCTGCTGCTGTTTGTGTTGCTTCCCT 24868
244 LysValPheLeuSerSerLeuLysPheHisIleArgArg 256
24869 TACCACGCTACTAAATCTAGACACTGCTGTAAGCGA 24907
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-466-103A-1
seq_documentation_block:
Sequence 1, Application US/08466103A
Patent No. 5856124
GENERAL INFORMATION:
APPLICANT: Reperet, Steven M.
APPLICANT: Ebisawa, Takashi
TITLE OF INVENTION: HIGH-AFFINITY MELANOTONIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,103A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/319,887

```



HYPOTHETICAL: NO  
NEXT-CASE: NO

```

FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-056-200-93

```

```

alignment_scores:
Quality: 106.00 Length: 277
Ratio: 0.774 Gaps: 7
Percent Similarity: 49.458 Percent Identity: 23.827

```

```

alignment_block:
US-09-510-332-1 x US-08-056-200-93/rev ..

```

```

Align seg 1/1 to reverse of: US-08-056-200-93 from: 1 to: 9551

```

```

2 LeuGluSerHisLeuIleIleTyPhe.....LeuGlyIlePheT 11
7347 CTGACGCGCGGAGCTGTTCTCTCCCTTCCTCGGAGCACTGTTCTTCA 7298
11 uLeuAlaValIleGluPheLeu.....LeuGlyIlePheT 23
7297 GGGGGGCGAGCTGTTCTCTCCCTTCCTCGGAGCACTGTTCTTCA 7248
23 hrasnglyIleIleValValaValasnGlyIleaspLeuIleIleYshArg 39
7247 CGGAATTTCTGCGGGCTGTGAGC.....CAGCTGTGTCGGCGCTC 7204
40 LysMetAlaProLeuaspLeuLeuLeuSerCysLeuAlaValSerArgI 56
7203 CTGGCGGCGGAGCTGCTCTCTCCCTCGAGGAATTTCTCTCAGGCTCT 7154
56 ePheLeuGluLeuPheIlePheTyPheValaValIleValIlePhePheI 73
7153 GGGCGGTGAGCTGTCTTCTTCCTTCCTTCGAGCACTGTTCTGTCTTCG 7104
73 LeuIlePheIleMetCysSerAlaasnGlyAlaIleLeuLeuPheIleasn 89
7103 CGGAATTTCTGCTCTCTCCCTCGGAGCTGTGTCGGCGCTCGGCGCG 7054
90 GluLeuGluLeuTrpLeuAlaIleTrpLeuGlyValPheTyPheCysAlaI 106
7053 CAGCTGCTGTCCTCCATGAATTTCTCTCTTCTTCCTCGGCGGCG.. 7006
106 sValAlaSerValArgHisProLeuPheIleTrpLeuLysMetArgIles 123
7005 .....CACTTCTGTCTCTCAACGGAATTTCTGTCA 6972
123 eTrpLysLeuValProTrpMetIleLeuGlySerLeuLeuValSerMet 139
6971 CG.....CTTGGGCGCTCAGCTGCTCTCTCTCTCTC 6937
140 IleCysValPheHisSerLysTyPheIleGluPheMetValProTyPheI 156
6936 CTGGCGGAGCTG.....TTCCTCTCT 6917
156 uArgLysPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrIle 173
6916 CGCGGAATTTCTGTACGCTCTCGCGCGCTGTGTCGCCGCGCTCTG 6867
173 lalIleGlnIlePheSerPheValAlaGluPheSerValProLeuLeuI 189
6866 CGGCGAGCTGCTCTCTCTCGGAGATTTCTCTGCGTTGCTGCGG 6817
190 PheLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThr 206
6816 GTGCACCTGCTGTCTCTCTCGAGGAATTTCTCTGTCTCTGCTGCTG 6767

```

```

206 gGlnMetArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaP 223
6766 GCAGTTCCTGTTCTCGGGAATTTCT..... 6742
223 roIleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyPheSer 239
6741 ..GTACGCTCTTGGCGCGGCGAGCTGTTCTCTCTCCATGAATTTCT 6694
240 HisCysMetIleLysValPheLeuSerSerLeuLysPheHisIleArg 256
6693 GTACGCTCTTGGCGGCTGACGCTGTTCTCTCTCTCTGAGCAGCT 6644
256 gPheIlePheLeuPhePheIleLeuValIle 266
6643 GTTCTCTCTCGGGAATTTCTGTGAGCTC 6613
seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-08-800-644-93
seq_documentation_block:
; Sequence 93, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichothyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fredrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1645..2511
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 2512..8070
; US-08-800-644-93

alignment_scores:
    Quality: 106.00      Length: 277
    Ratio: 0.774         Gaps: 7
    Percent Similarity: 49.458    Percent Identity: 23.827

alignment_block:
    US-09-510-332-1 x US-08-800-644-93/rev ..

Align seg 1/1 to reverse of: US-08-800-644-93 from: 1 to: 9551

2 LeuGluserHisLeuIleIleTyPhe.....Le 11
   ||| :|||:|||||:|||||:|||||:|||||:
7347 CTCACGGCGGAGCTGCTGCTTCCTCCGCGGAATTTCTGTACGCTCTT 7298
   ||| :|||:|||||:|||||:|||||:|||||:
11 uLeuAlaValIleGlnPheLeu.....LeuGlyIlePheT 23
   ||| :|||:|||||:|||||:|||||:|||||:
7297 GCGCGGCGAGCTGCTGCTCTCCCTCTGAGCAGCTTCCTCTCA 7248
   ||| :|||:|||||:|||||:|||||:|||||:
23 hTAsnGlyIleIleValIleValAsnGlyIleAspLeuIleYHisArg 39
   ||| :|||:|||||:|||||:|||||:|||||:
7247 CCGAATTTCTGTCCGCGCTGAGC.....CAGCTGTGTTCGCGCTC 7204
   ||| :|||:|||||:|||||:|||||:|||||:
40 LysMetAlaProLeuAspLeuLeuSerCysLeuAlaValSerArg11 56
   ||| :|||:|||||:|||||:|||||:|||||:
7203 CTCGCGCGCGACCTCGGCTTCCTCGAGAAATTTCTGTACGCTCTT 7154
   ||| :|||:|||||:|||||:|||||:|||||:
56 ePheLeuGlnLeuPheIlePheTyValAsnValIleValIlePhePheT 73
   ||| :|||:|||||:|||||:|||||:|||||:
7153 GCGCGGCGAGCTGCTGCTTCCTCTTCCTGAGCAGCTGTGCTTCG 7104
   ||| :|||:|||||:|||||:|||||:|||||:
73 LeuGlnPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsn 89
   ||| :|||:|||||:|||||:|||||:|||||:
7103 CCGAATTTCTGTCTCCGCGGAGCTGTGCTGCGCTCTGCGCGCG 7054
   ||| :|||:|||||:|||||:|||||:|||||:
90 GluLeuGlnLeuTrpLeuAlaThrTrpLeuGlyValPheTyrcysAlaLy 106
   ||| :|||:|||||:|||||:|||||:|||||:
7053 CAGCTGCTGCTGCTCTCCAGAAATTTCTGTCTTCCTGCGCGCG.. 7006
   ||| :|||:|||||:|||||:|||||:|||||:
106 sValAlaSerValArgHisProLeuPheIleTrpLeuYsMetArg11es 123
   ||| :|||:|||||:|||||:|||||:|||||:
7005 .....SACTTCTGTTCTCTTAACGGAATTTCTGTCTGCA 6972
   ||| :|||:|||||:|||||:|||||:|||||:
123 eLysLeuValProTrpMetIleLeuGlySerLeuLeuTyrcysValSerMet 139
   ||| :|||:|||||:|||||:|||||:|||||:
6971 CG.....CTTTCGCGGCTCAGCTGCTGCTCTCCCTCCTC 6937
   ||| :|||:|||||:|||||:|||||:|||||:
140 IleCysValPheHisSerLysTyrcysAlaGlyPheMetValProTyrcPheLe 156
   ||| :|||:|||||:|||||:|||||:|||||:
6936 CTCGCGGCGAGCTC.....TTTCCTCTC 6917
   ||| :|||:|||||:|||||:|||||:|||||:
156 uArgLysPhePheSerGlnAsnAlaThrIleGlnLysGlnAspThrLeuAla 173
   ||| :|||:|||||:|||||:|||||:|||||:
6916 CCGGGAATTTCTGTACGCTCAGCGCGCTGTCGCGCGCTCTCTCG 6867
   ||| :|||:|||||:|||||:|||||:|||||:
173 IaIleGlnIlePheSerPheValAlaGlnPheSerValProLeuLeuIle 189
   ||| :|||:|||||:|||||:|||||:|||||:
6866 CCGCGCGAGCTGCTGCTCTCCGAGGAATTTCTTCGCTGCTGCGCG 6817
   ||| :|||:|||||:|||||:|||||:|||||:
190 PheLeuPheAlaValLeuLeuIlePheSerLeuGlyArgHisThrAr 206
   ||| :|||:|||||:|||||:|||||:|||||:
6816 GTGAGCTGCTGCTCTCCTCGAGGAATTTCTCTGCTGCTCTGAGCTGC 6767
   ||| :|||:|||||:|||||:|||||:|||||:
206 gGlnMetArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaP 223
   ||| :|||:|||||:|||||:|||||:|||||:
6766 GCAGTTCCTGTCGCGGAATTTCT..... 6742
   ||| :|||:|||||:|||||:|||||:|||||:
223 rolIeSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyrcPheSer 239
   ||| :|||:|||||:|||||:|||||:|||||:

```

```

6741 ..GTCACGCTCTTGGCGGCGAGCTGTGCTCTCCAGGAATTTCT 6694
240 HisCysMetIleLysValPheLeuSerSerLeuLysPheHisIleArgAr 256
   ||| :|||:|||||:|||||:|||||:|||||:
6693 CTCACGCTCTTGGCGGCTGCTGCTCTTCCTCCCTTCCTGAGCAGCT 6644
   ||| :|||:|||||:|||||:|||||:|||||:
256 gPheIlePheLeuPhePheIleLeuValIle 266
   ||| :|||:|||||:|||||:|||||:|||||:
6643 GTTCCTCTTCGCGGAATTTCTGTGAGCCTC 6613

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-137-175A-7

seq_documentation_block:
; Sequence 7, Application US/08137175A
; Patent No. 5777095
; GENERAL INFORMATION:
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: BERGSTROM, Sven
; APPLICANT: HANSSON, Lennart
; TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND
; TITLE OF INVENTION: PROPHYLAXIS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NETMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/137, 175A
; FILING DATE: 26-OCT-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08972
; FILING DATE: 22-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28, 005
; REFERENCE/DOCKET NUMBER: BARBOUR-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: IP90
; INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
; INDIVIDUAL ISOLATE: Soviet Union
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 125..949
; OTHER INFORMATION: /product= "Ospa"
; NAME/KEY: CDS
; LOCATION: 959..1843
; OTHER INFORMATION: /product= "Ospb"
; US-08-137-175A-7

alignment_scores:

```

Quality: 104.50 Length: 349  
Ratio: 0.611 Gaps: 21  
Percent Similarity: 48.997 Percent Identity: 23.496

Alignment\_block:  
US-09-510-332-1 x US-08-137-175A-7/rev ..

Align seg 1/1 to reverse of: US-08-137-175A-7 from: 1 to: 1958

```

1 MetLeuGluSerHisLeuIleLeuPheLeuLeuAlaValIle..... 15
   |||||.....
1034 GTTTGGCTGACGACCTTTTGTGTCACATGCTATTAAAGCTAATACTATA 985
   |||||.....
16 GluPheLeuGluGlyIlePheThrAsnGlyIleLeuValValAsnG 32
   |||||.....
984 GCAATTCCTAGTAATATTTTTC.....TAA 956
   |||||.....
32 LysIleAspLeuIleLysHis.....Arg 39
   |||||.....
955 CTCTCCCTATTTTTAAAGCATCTTTAAGTCTTTAAGCGCTGTAATTTGCA 906
   |||||.....
40 LysMetAlaProLeuAspLeuLeuSerCysLeuAlaValSerArg11 56
   |||||.....
905 CTGCTTGGCTCTAGATTGG.....TGCCCTG..CTGAGTCGAT 868
   |||||.....
56 ePhe.....LeuGluLeuPheIlePheT 64
   |||||.....
867 TTTTGTACTGTTATTGCTCTTTTGTGTAATACAGAGTTTGGT 818
   |||||.....
64 yr.....ValAsnValIleVal 69
   |||||.....
817 TCGGCTATTACACATAATTGTTAAAGTGAAGCTTTGATCCAGTTC 768
   |||||.....
70 IlePhePheIleGluPheIleMetCysSerAlaAsnCysAlaIleLeu 86
   |||||.....
767 CAGTTTTTTACAGCTGAGTAGTGTACAGATTAAGCTCAACTGTT 718
   |||||.....
86 upheIleAsnGluLeuGluLeuTLeuAlaThrTLeuGlyValPheT 103
   |||||.....
717 ATTTCTCCAGAGTTGAAATGT.....G 695
   |||||.....
103 yrCysAlaLysValAlaSerValArgHisProLeuPheIleThrLeuLys 119
   |||||.....
694 TTGGCTTAAACAAACAGTCCCTCTGTAATTTAATGTTTGGCCG 645
   |||||.....
120 MetArgIleSerLysLeuValProThrMetIleLeu..GlySerLeuLeu 135
   |||||.....
644 CACGAGCTAGAGTTCCCTTCAAGCAAGCAAGCTTTTAAACTCTTAGCT 595
   |||||.....
136 TyrValSerMetIleCysValPheHisSerLysTyrAlaGlyPheMetVal 152
   |||||.....
594 TTT...CCGGTTTATCGCTTTTATGCTGTGATTTCAAGT...CTGGT 551
   |||||.....
152 LProTyrPheLeuArgLysPhePheSerGlnAsnAla.....T 165
   |||||.....
550 TCCATTGCTCTTACTATTGTTTTCAGATGCTTCACCTTGGCGTTGA 501
   |||||.....
165 hrIleGlnLysGluAspThrLeuAlaIleGlnIlePheSerPheValAla 181
   |||||.....
500 ATTTTCTTCTGTGATGATGCTGCTTAAAGGTTACTTTTGTGATACT 451
   |||||.....
182 GluPhe.....SerValProLeuLeu..... 188
   |||||.....
450 AATGTTTGGCATCTCTTGAAGATTTCAATATGTTTACTTAATC 401
   |||||.....
189 .....IlePheLeuPheAlaValLeuLeuIlePheSerLeuGly 203
   |||||.....
400 CTCAGCATTTGTTAATTTGCTTACTTTTCTCAGTTTTC..... 358
   |||||.....
203 rghIstHrArgGlnMetArgAsnHrValAlaGlySerArgValProGly 219
   |||||.....
357 .....CCTTCAAGTGTTCAGAA 340

```

```

220 .....ArgGlyAlaProIleSerAlaLeuSerIleLeuSe 232
   |||||.....
339 CCCTGTTTATTACAGAGTTCCTTTAAGCTCAAGCTTGTACTGTTCC 290
   |||||.....
232 rPheLeuIle..LeuTyrPheSerHisCysMetIleLysValPheLeuSer 248
   |||||.....
289 CATTAGACTGATTTAC.....CATCTTGTCTTTTCTTACTTACAA 246
   |||||.....
249 SerLeuLysPheHis.....IleArgArgPheIlePhe..... 259
   |||||.....
245 GAACCTTCATTTCCACGAGTAATCTACTGAAACGCTATTTTTCA 196
   |||||.....
260 .....LeuPhePheIleLeuValIleGlyIleT 269
   |||||.....
195 AGCGTGTACATTTCCTTACATGCTATTATGCTAATTTAGACTAT 146
   |||||.....
269 yrPro.....SerGlyHisSerLeuIleLeuIle 279
   |||||.....
145 TCCCAATTAATATTTTTCATTAATATTTCTCCTTTTATATTA 102
   |||||.....
seq_name: /cgn2_6/pdata/2/ina/6_COMB.seq:us-08-479-017-7
seq_documentation_block:
; Sequence 7, Application US/08479017
; Patent No. 6143872
; GENERAL INFORMATION:
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: BERGSTROM, Sven
; TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDOFFERI AND
; TITLE OF INVENTION: PROPHYLAXIS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,017
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/137,175
; FILING DATE: 26-OCT-1993
; APPLICATION NUMBER: PCT/US92/08972
; FILING DATE: 22-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: BARBOUR-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 246633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: Ip90
; INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
; INDIVIDUAL ISOLATE: Soviet Union

```



```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-487-826B-13

alignment_scores:
  Quality: 104.00      Length: 331
  Ratio: 0.588         Gaps: 15
  Percent Similarity: 53.474   Percent Identity: 21.148

alignment_block:
  US-09-510-332-1 x US-08-487-826B-13

Align seg 1/1 to: US-08-487-826B-13 from: 1 to: 19124

1 MetLeuGluSerHisLeuIleIleTyrrPheLeuLeuAlaValIleGlnPhe 17
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17446 AGCTTAAATAATCATCATGATGATGATATATATGATGATCATATATATG 17495
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17   eLeuLeuGlyIlePheThrAsnGlyIleIleValValValAsnGlyIleA 34
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17496 ACTAATAATAATGATGATGATGATGATGATGATGATGATGATGATGAT 17545
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 spleuIleLysHisArgLysMetAlaPheLeuAspLeuLeuSerLysCys 50
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17546 TGTAAATAATAATAAAGAGATTTTCGACAGAGATATCTCATATATGATA 17595
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 LeuAlaValSerArgIle.....PheLeu.....GlnLeu P 61
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17596 TAGGAATATCTAAATAATATATATCTTTTGTGTGTCATATATATATAT 17645
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 heilePheTyrrValAsnValIleValIlePhePheIleGluPheIleMet 77
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17646 TTGATATATTTGATATATGTTTATTTTATTTATTTATTTATTTATTTAT 17695
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
78 CysSer...AlaAsnCysAlaIleLeuLeuPheIle...AsnGluLeuG 92
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17696 TGTTTTGGTATATTTGTAAAAAATATGTTTGTATATATATATATATAT 17745
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 uLeuTyrrLeuAlaThrTyrrLeuGlyValPheTyrrCysAlaLysValAla 109
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17746 TATATTTTATATATTTGCAACATGATTTTCTTTCTTTCTTTATTTGTC 17795
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 eValAlaGHisProLeuPheIleTyrrLeuLysMetArgLysSerLysLeu 125
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17796 TAAATTTTTCATATAATTTATATATATATATATATATATATATATAT 17830
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 ValProIrrMetIleLeuGlySerLeuLeuTyrrValSerMetIle..... 140
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17831 .....TTTATTTTATATATATATATATATATATATATATATATATAT 17874
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 CysValPheHisSerLysTyrrAlaGlyPheMetValProTyrrPheLeuA 157
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17875 TTATCTAATATATATGTTATTTGTTCTTTCTTTCTGACATTTTTCGAA 17924
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 rGlyPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeuAla 173
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17925 TATATATATATAT.....ATATATATATATATATATATATATATATAT 17959
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 IleGlnIlePheSerPheValAlaGluPhe.....SerValProLeuLe 188
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17960 CTAAATATATATATCTCTATTTTCGATTTTTCATTTTTCCTGAGTAT 18009
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
188 uIlePheLeuPhe..... 192
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18010 AATTTATTTATTTATTTGATATTTTATATATATATATATATATATATG 18059
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 .....AlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArg 206
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

18060 TTTATATATGCTGTTTATATATGCTGTTTATTTT.....GTACT 18100
207 GlnMetArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaP 223
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18101 CTAAATCTGAATATATCCGACGCAAAAAAATATATATATATCTC..... 18142
223 oIleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyrr..... 237
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18143 .ATATATAAATTTATTTATATATATATATATATATATATATATATATAT 18191
238 .....PheSerHisCysMetIleLysValPheLeuSerSerLeu 250
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18192 AATTAATATATATATACAAATATATATTTCTGTTATTTTATATAATATAC 18241
251 LysPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18242 AATTTCTTATTTTATTTTAACTTATTCCTTTTAAATTTCTTATATCTT 18290
267 LylleTyrrProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProLys 283
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18291 ..TTATCAACAAAAACATAATATATCTCATATATCAACAAAAA 18338
284 LeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLys 296
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18339 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 18377
seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-748-506-9
seq_documentation_block:
  Sequence 9, Application US/08748506
  Patent No. 6159707
  GENERAL INFORMATION:
  APPLICANT: Ronnett et al.
  TITLE OF INVENTION: NOVEL SPERM RECEPTORS
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Leydig, Volt & Mayer, Ltd.
  STREET: Two Prudential Plaza, Suite 4900
  CITY: Chicago
  STATE: IL
  COUNTRY: US
  ZIP: 60601-6780
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/748,506
  FILING DATE: 08-NOV-1996
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 60/033,751
  FILING DATE: 09-NOV-1995
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  REFERENCE/DOCKET NUMBER: 74940
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 312-616-5600
  TELEFAX: 312-616-5700
  INFORMATION FOR SEQ ID NO: 9:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 984 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  US-08-748-506-9

alignment_scores:
  Quality: 103.50      Length: 265
  Ratio: 0.690         Gaps: 12

```



420 ..... 420  
74 uphelleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnIuL 91  
421 .....GCTGACCTGACCTTCGTGGTGACGC 445  
91 euGIuLeuTriPheuAlaThrTrp.....Leu 99  
446 TCCCCCTGTGG...GCTACCTACAGTACCGGACTAGACTGGCCCTTT 492  
100 GlyValPheTYrCys..... 104  
493 GGGACCTTCTTCGACAGCTCAGCAGTACCTCATCTTCGTCAACATGTA 542  
105 .....AlaLysValAlas 109  
543 CCGCAGCGTCTTCCTCACCAGCGCTCAGCTTCGACCGCTACCTGGCCA 592  
109 exValArgHisProLeuPheIleTrPLeuLysMetArgIleSerLysLeu 125  
593 TCGTGAAGCCATGGCCATGCTCGGCTGAGGCTGGGGTCAAGCGGGCC 642  
126 ValProTrpMetIleLeuGlySerLeuLeuTYrValSerMetIleCysVa 142  
643 GTGGCCACGCGAGTCTGTGGGTGGCGCGCCCTTCCTGGCCATGCTGT 692  
142 IPhenIleSerLysTYrAlaGlyPheMet.....ValProTY 154  
693 CATGTGTTTACGACCAACCGGACTTGGAGAACACCACTAAGTGTCAGT 742  
154 YrPheLeuArgLysPhePheSerGlnAsnAlaThrIleGlnLysGluAsp 170  
743 GCACATGGAC.....TACTCATGTGGCGCCACTGTGAGCTCAGAGTGG 786  
171 Thr.....LeuAlaIleGlnIlePheSerPheValAlaGluPheSerVa 185  
787 GCCTGGAGGTGGGCTTGGGTGCTGCACCAACCGTGGCTTTGTGGT 836  
185 IProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSerLeuG 202  
837 GCCCTTACCATCATGCTGACCTGTACTCTTCATCGCCCAACCATCG 886  
202 IyArgHisThrArgGlnMetArgAsnThrValAlaGlySerArgValPro 218  
887 CTGGCCACTTCCGCAGAGAACGC.....ATCGAGGGCTCGGAGACGG 930  
219 GLyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheLeuI 235  
931 CGCCGG.....CTGCTCAGCATCATCGTGGTGTGGT 962  
235 eleuTYrPheSerHisCysMetIleLysValPheLeuSerSerLeuLysP 252  
963 GGTGACCTTTCCTGCTGTGG.....ATGCCCT 991  
252 heHisIleArgArgPheIlePheLeu..... 260  
992 ACCACCTGTGTAAGAGCTGTACATGTGGGACGACCTGTGCTGAGGCC 1041  
261 .....PhePheIleLeuValIleGlyIleTYrPro..... 270  
1042 TGTGACTTTGACCTCTTCATGACATCTTCGCTACCTGACCTGCAT 1091  
271 .....SerGlyHisSerLeuIleLeuIleLeuGlyAsnP 282  
1092 CAGTACGTCAACAGCTGCTCAACGCTTCCTATGCTTTTTCGAC 1141  
282 polyLysLeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCys 298  
1142 CCGGCTTCGCGGAGGCTGACCTGCATGCTC.....TGTGT 1179

**THIS PAGE BLANK (USPTO)**







100

33

```
FT /note="Substitution with A in mutant sequence"  
FT 362..381  
FT /*tag= h
```



[illegible]

seq_name:	/cgn2_2/gcgdata/geneseq/geneseq/NM1997.DAT:T69547
seq_documentation_block:	
ID	T69547 standard; DNA; 1496 BP.
XX	
AC	T69547;
XX	
DT	26-AUG-1997 (first entry)
XX	
DE	Rat pheromone receptor VN4 CDNA.
XX	
KM	Pheromone receptor; vomeronasal sensory neuron; social behaviour.
KW	Maternal behaviour; reproductive behaviour; fertility;
KW	hormone secretion; ss.
XX	
OS	Rattus sp.
XX	
FH	Key
FT	Location/Qualifiers
FT	351..1283
PN	/tag= a
XX	
PN	WO9714790-A1.
XX	





```

KM hereditary; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH prim_transcript 200..1341
FT /*tag= a
FT CDS 295..1341
FT /*tag= b
XX
XX MO9312134-A.
XX PD 24-JUN-1993.
XX
XX 08-DEC-1992; 92MO-US10536.
XX
XX 11-DEC-1991; 91US-0805123.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Berson EL, Dryja TP;
XX WPI: 1993-214088/26.
XX P-PSDB: R38483.
XX
XX Probe or primer conty. sequence of human retinal degeneration
XX slow protein mutant - used to diagnose hereditary retinal
XX degenerative diseases
XX
XX Disclosure: fig 1; 56pp; English.
XX
XX The sequence given represents the human rhodopsin cDNA. Mutant
XX versions of this sequence encode proteins which cause retinal
XX degeneration. These sequences may be identified using primers/
XX probes described in the invention (see also 043545-48) and may be
XX used to diagnose hereditary retinal degeneration. This sequence is
XX the closest approximation to the gene sequence as the sequence given
XX in the specification is not printed clearly.
XX
XX Sequence 3129 BP; 694 A; 999 C; 654 G; 735 T; 47 other;
XX
XX
XX alignment_scores:
XX Quality: 122.50 Length: 327
XX Ratio: 0.756 Gaps: 16
XX Percent Similarity: 49.541 Percent Identity: 21.101
XX
XX alignment_block:
XX US-09-510-332-1 x 043543 ..
XX
XX Align seg 1/1 to: 043543 from: 1 to: 3129
XX
XX 1 MetleuGuseRHISleuIleIleTyPheleuLeuAlaValIleGlnP 17
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 409 ATGCTGGCCGCTACATG.....TTTCTGTGATCGTCGCGGCTT 449
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 17 eleuLeuGlyIlePheRhrAsnGlyIleIleValValAlaSnGlyIleA 34
XX | :|||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 450 CCCCATCAACTTCCTCACGCTCTACGTCACCGCTC..... 483
XX
XX 34 splleuIleuYHisArgLysMet...AlaProLeuAspLeuLeuSer 49
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 484 .....CAGCAACAAGAGCTGCGCACCCTCTCACTACATCTCTGTC 525
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 50 CysLeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyRValAs 66
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 526 AACCTAGCGCTGCTGACCTCTTCATGCTAGCTAGCTGCTCACCAGCAC 575
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 66 nValIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysA 83
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 576 CCTTACACCTCTCTGCATGATGATCTGCTTGGGGCCACAGGATGCA 625
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 83 laIleLeuPhePheIle.....AsnGluLeuGlnLeuTyR... 94

```

```

626 ATTTGGARGGCTTCTTTGCCACCCTGGCGGTGAATATGTCCTGGTCC 675
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 95 .....LeuAlaThrTyPheGlyValPheTyRysAlaValAl 108
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 676 TTGGTGCTCGTGGCATCGACGCTAGCTGTGTGTAAACCCATGAG 725
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 108 aserValArg.....HisProLeu.....PheIleT 117
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 726 CAACCTCCGCTTCGGGAGAGAACCATCATCATAGCGCTTCACT 775
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 117 rp.....LeuYMetLArgIleSerLysLeuValProTyPmetIle 130
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 776 GGTTCATGGCGCTGCGCTGCGCCGACCCCACTCGCGGCTGG..... 819
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 131 LeuGlySerLeuTyRValSerMetIleCysValPheHisSerLysTy 147
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 820 .....TCCAGGTA 827
XX
XX 147 rAlaGlyPheMetValProTyRPhleuArg.....LysP 159
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 828 C.....ATCCCGAGGCGCTGCACTGCTCGTGGAAATGCACT 865
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 159 hepSerGlnAsnAlaThrIleGlnLysGluAspThrLeuAlaIleGln 175
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 866 ACTTACACCTTCAGCGCGAGGTCAACACGAG.....TCTTTGTC 906
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 176 IlePheSerPheValAlaGluPheSerValProLeuLeuIlePheLeuP 192
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 907 ATCTACATGTTCTGTGCTCCTTCCACATCCCATGATATCATCTTTT 956
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 192 eAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArgGlnMet 209
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 957 CTGCTATGGGCAAGCTGCTTTCACCGTCAAGAGGCGCGCCCAACAGC 1006
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 209 rGAsnThrValAlaGlySerArgValProGlyArgGlyAlaProIleSer 225
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1007 AGGAGTCA...GCCACACACAGAGCAGAGAMGNNNGTCACCGCATG 1053
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 226 AlaLeuLeuSerIleLeuSerPheLeuIleTyRPheserHisCysMe 242
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1054 GTCATCATCATGTCATCGCTTCTGATCTCTGGTCCCTACGCCAG 1103
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 242 tIleLysValPheLeuSerSerLeuLysPheHisIleArgArgPheIleP 259
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1104 CGTGCATCTTACATCTTCACCCACAGGCTCAACTGCTGCCACT 1153
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 259 heLeu.....PhePheIleLeuValIleGlyIleTyRProSer 271
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1154 TCATGACCATCCACGCGCTCTTGGCCAACTCGCGCCCATCTCAACCT 1203
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 272 GlyHisSerIleLeuLeuGlyAsnProLysLeuLysGlnAsnAl 288
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1204 GTCATCATATCATGATGATC.....AACTGTCCGGAGACG 1241
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 288 aLysLysPheLeuLeuHisSerLysCysCys 298
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1242 C.....ATGCTACCCACCATCTGCTGC 1263
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX seq_name: /cgn2.2/gcdata/geneseq/geneseq/NM1998.DAT:V31718
XX
XX seq_documentation_block:
XX ID V31718 standard; DNA; 1202 BP.
XX
XX AC V31718;
XX
XX XX
XX DT 11-SEP-1998 (first entry)
XX
XX DE Kaposi's sarcoma associated herpesvirus ORF74.
XX
XX XX
XX KW PCR primer: KSHV; ORF74; Kaposi's sarcoma; diagnosis; treatment;
XX G protein-coupled receptor; ss.
XX
XX

```





```

FT CDS 90173..90643
FT FT /*tag= k
FT FT /product= glycoprotein X
FT CDS complement (93636..94127)
FT FT /*tag= 1
FT FT /product= interferon regulatory factor 4
FT CDS complement (11931..112443)
FT FT /*tag= m
FT FT /product= capsid protein IV
FT CDS complement (123808..127296)
FT FT /*tag= n
FT FT /product= immediate early protein

W09804576-A1.
PD 05-FEB-1998.
XX 22-JUL-1997; 97WO-US13346.
XX 29-NOV-1996; 96US-0757669.
XX 25-JUL-1996; 96US-0686243.
XX 25-JUL-1996; 96US-0686349.
XX 25-JUL-1996; 96US-0686350.
XX 25-JUL-1996; 96US-0687253.
XX 25-JUL-1996; 96US-0688814.
XX 05-SEP-1996; 96US-0708678.
XX 10-OCT-1996; 96US-0728323.
XX 13-NOV-1996; 96US-0747887.
XX 13-NOV-1996; 96US-0748640.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
XX WPI: 1998-130615/12.
XX
PT New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT proteins - useful for e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients
PT
PS Example 2; Page 135-203; 230pp; English.
PS
XX This sequence represents the long unique region and terminal repeat of
XX the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
XX as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
XX invention which encode KSHV polypeptides selected from: (a) viral
XX macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
XX (c) viral INF 1; (d) complement-binding protein; glycoproteins B, M or L;
XX (e) capsid protein IV encoded by ORF5; and (e) immediate early protein
XX encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
XX by it, and antibodies (Ab) specific for the proteins are useful for
XX detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
XX fluids or tissue samples. HHV8 infections can be treated with antisense
XX or triplex forming molecules or agents that bind specifically to the
XX protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
XX to differentiate between lymphomas, and HHV8 may be implicated in many
XX other lymphoproliferative diseases such as lymphomas, leukaemia,
XX splenomegaly and mycosis fungoides. Cells and animals containing the
XX nucleic acid are useful for drug screening. HHV8-derived peptides can be
XX used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
XX can be inhibited with methotrexate. These can also be used to determine
XX the immune status of a patient infected with HIV. HHV8 derived protein
XX viral MIP III may be used as an anti-inflammatory agent for,
XX e.g. treating rheumatoid arthritis. This sequence is stated as containing
XX 81 open reading frames.
XX
Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

```

Alignment\_scores: Quality: 120.50 Length: 263  
Ratio: 0.867 Gaps: 14

```

Percent similarity: 52.852 Percent identity: 22.433
Alignment block:
US-09-510-332-1 x V19941
Align seg 1/1 to: V19941 from: 1 to: 137507

11 LeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThrPheArgIleIleI 27
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
129527 ATACTCTCTCGATT...TTCTCATTAATGTTCTTGGAAATGATTCGT 129573
27 eValValValAsnGlyIleAspLeuIleLysHisArgLysMetAla...P 43
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129574 CACCTACATT...TTTGCACACCGATCCGATCCGCGCAGAG 129611
43 rOlAuAspLeuLeuLeuSerCysLeuAlaValSerArgIlePheLeuGln 59
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129612 CGATAGATATATCTCTCTCGTGGTATCTGCTTAACTGCTGTCTTATGC 129661
60 LeuPheIlePheThrValAlaValIleValIlePhePheIleGluPheI 76
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129662 ATATCTCTA...TTGCCAGAGTGTGATGATTTTGTTCCTCAATAT 129705
76 eMeCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnGluLeuGlu 93
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129706 CATCTCCACAGGCTTGTGCGACACTTGAAATTTT...T 129743
93 eUrTrpLeuAlaThrTrpLeuGlyValAlaPheThrCysAlaLysValAlaSer 109
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129744 ACTATTTATATGTTCTACTGATATCTTCAAGTGTGTCGTCACATCTA 129793
110 ValArgHisProLeuPheIle...TrpLeuLysMetArg 121
|||||:::|||||:::|||||:::|||||:::|||||:::
129794 GTGAGTAACTCTCTCTGTCGGGATCTTACCGCTTCTCGGCCAAG... 129838
121 gIleSerLysLeuValProTrpMetIleLeuGlySerLeuLeuTrpValS 138
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129839 ...AAGCAGTCCCTCGATGGGTACTGACATCCGCTGACATCTTAATG 129884
138 eMeTlle...CysValPheHisSerLysTrpAlaGly 149
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129885 CATTTGGCTGTGCGGGGATCCCTGTGCACACGACGAGCGGTGTCAG 129934
150 PheMetValProTrpThrPheLeuArgLysPhePheSerGlnAsnAlaThrI 166
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
129935 .....CCGTCAGCAAGCAGCGCATCTGTTATGACACAGCGGGA 129975
166 eGlnLysGluAsp...ThrLeuAlaIleGlnIlePheSerPheValAlaG 182
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129976 CATTGACTGCAGACTGCGACTGCATGTCAGAACCCGTGATGTTACTG 130025
182 IupHeserValProLeuLeuIlePheLeuPheAlaValLeuLeuLeu 198
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
130026 GTTCTCTGTACCCCTG.....GCCCTCTTATTCGTTT 130060
199 PheSerLeuGly.....ArgHisThrArg...GlnMetArgAs 210
:::|||||:::|||||:::|||||:::|||||:::|||||:::
130061 TATGCTCTCACTGTGTGTGTGAGAGCAGACAAAGCTGACCCAGGGC 130110
210 nTrpValAlaGlySerArgValProGlyArgGlyAlaProIleSerAla 227
:::|||||:::|||||:::|||||:::|||||:::|||||:::
130111 GAAGGTAAAGGGG..... 130123
227 eUleuSerIleLeuSerPheLeuIleLeuTrpPheSerHisCysMetIle 243
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
130124 ....GTGATGTGTGCTGTGCTGTGCTGTTTGTGTTGCTTCCCT 130168
244 LysValPheLeuSerSerLeuLysPheHisIleArgArg 256
|||||:::|||||:::|||||:::|||||:::|||||:::
130169 TACCACTACTAATCTACTGACACTGTGCTAAGCGCA 130207
seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT.T79064

```





```

XX AC T79063;
XX DT 10-NOV-1997 (first entry)
XX DE Xenopus melatonin receptor MEL-1Aa long form coding sequence.
XX XX
XX Allele: Xenopus laevis; melatonin; receptor; untranslated region; PCR;
XX KW mRNA; half-life; skin; amplification; primer; polymerase chain reaction;
XX KW transmembrane domain; cellular signalling; inhibition; adenylyl cyclase;
XX KM modulation; intracellular; cyclic GMP; inhibitor; phosphodiesterase; ss.
XX OS Xenopus laevis.
XX FH Key Location/Qualifiers
XX FT CDS 1..1065
XX FT /*tag= a
XX FT /product= MEL-1Aa receptor protein
XX PN W09704094-A1.
XX PD 06-FEB-1997.
XX PE 24-JUL-1996; 96MO-FR01167.
XX PR 24-JUL-1995; 95FR-0008947.
XX PA (ADIR ) ADIR & CIE.
XX PI Jockers R, Marullo S, Strosberg AD;
XX DR WPI: 1997-132635/12.
XX DR P-PSDB; W25926.
XX PT New nucleic acid encoding functional melatonin receptor of Xenopus -
XX PT for screening for potential (ant)agonists useful for e.g. treating
XX PS cardiovascular disease and cancer
XX PS Claim 1; Page 25-26; 62pp; French.
XX
XX Sequences T79063-66 represent novel allelic genes of the Xenopus laevis
XX melatonin receptor MEL-1a. The sequences encode proteins which are 65
XX amino acids shorter than those described in the prior art. Also the last
XX 2 C-terminal amino acids encoded by these sequences are different from
XX the previously known proteins. This sequence is a long form of the novel
XX receptor MEL-1Aa also known as Mel 1-C(alpha). As compared to the short
XX form (T79064), the difference occurs in the 3' untranslated region (both
XX sequences encode the same protein), which is thought to affect the
XX half-life of the mRNA. The nucleotide sequence was isolated from cDNA
XX derived from Xenopus skin RNA and amplified using the primers T79067-76.
XX The nucleotide sequence encodes a protein which is a 7 transmembrane
XX receptor involved in cellular signalling. MEL-1Aa has been shown to
XX inhibit adenylyl cyclase, but both proteins can modulate intracellular
XX cGMP, esp. inhibiting its accumulation induced by an inhibitor of
XX phosphodiesterase.
XX
XX Sequence 1311 BP; 369 A; 259 C; 269 G; 414 T; 0 other;
XX
XX alignment_scores:
XX Quality: 118.00 Length: 294
XX Ratio: 0.752 Gaps: 12
XX Percent Similarity: 53.401 Percent Identity: 19.388
XX
XX alignment_block:
XX US-09-510-332-1 x T79063 ..
XX
XX Align seg 1/1 to: T79063 from: 1 to: 1311
XX
2 LeuclSerHisIeuLeuIleTyrPheLeuLeuAlaValIleGlnPheLe 18
|||||
1 CTGACCTGCCTGCAGGAGTGTCTTATATATTCACCATTTGTGTGATGT 140

```

18 uLeuGlytIlePheThrAsnGlyIleIleValValAlaAsnGlyIleAspL 35  
19 |||||  
141 CCGTGGC..... AATAATTGGTCATT..... TTGTCTG 169  
35 euIleIySHisArgMetAlaProLeu... AsPLeuLeuSerCys 50  
170 TCGGAGACAACAAGACCTGCAGAAATCGTGGAAATCTCTTGTGTGCAGT 219  
171 |||||  
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyValAsnVa 67  
220 TTGTCTATTCGCCGATCTGTGGTTGTGTGTGTATCCCATCCGCTAATTC 269  
67 |||||  
67 ILeIleAlIlePhePheIleGluPheIleMetCysSerAlaAsnGlyAlaI 84  
270 CATACCTATTTTCCAGAAATGGGTGCAGCGCTTGGAATAATCATTCATG 319  
84 |||||  
84 LeuLeuPheIleAsnGluLeuGluLeuTyrPLeuAlaIleTyrPLeuGly 100  
320 TCAGTGGCTTCTCGATGGACCTACCGTT..... ATTGGA 354  
101 ValPheTyCysAlaIleValAlaSerValArgHisProLeuPheIleTr 117  
355 TCAGCTTTCACATAAACAGCCATACCTATCTTCAACAAGATTTGCTACATCG 404  
117 |||||  
117 PleuTySMetArgIleSerIySLeuVal..... ProTyrPheMet 130  
405 CCACAGCCCGAGATATGACAAAGCTTTATATATCAAGAAGACACTGGTCT 454  
130 ILeuGlySerLeuLeuTyValSerMetIleCysValPheHisSerIys 146  
455 ACCCTGGCCGTGACATGATACATATATATGACATGTCGCCAAACTTT 504  
147 TTYcAlaGlyPheMetVal..... ProTyrPheLeuArgIysPhePheSe 161  
505 TTTGTTGGATCACTACAGATATGACCCCGACAGATTTTCTTCGACACTTGGC 554  
161 rGlnAsnAlaThrIleGlnIlySLeuAlaPheIleuAlaIleGlnIlePheS 178  
555 GCAG..... ACAGTACAGTTCCTCAATACACATACAGATGATGGTGTCC 598  
178 eArPheValAlaGluPheSerValProLeuLeuIlePhePheLeuAlaVal 194  
599 ATTTTATATAGTCCTCTTAGTGTTGTGCATCTGTCATCTTA... AGAATA 645  
195 LeuLeuLeuIlePheSerLeuGlyArgHisThrArgLn..... 207  
646 TGGATTATAGTATCCAAAGTCAACACAGAGTGTAGCAACAAGACTTCACGA 695  
208 ..... MetArgAsnThrValAla..... 213  
696 AAAGTTGACACAAACAGACTGTGAGAATTTCTTGACCATGTGTGGTGCT 745  
214 ..... GlySerArg 216  
746 TTGTACTTTTGCAGTTTGTGCGCCCTTAAACTTATACGGCGCTTGGCT 795  
217 Val..... ProGlyArgGlyAlaPro... IleSerAlaLeuLeuSe 229  
796 GTGGCCATTAAATCCGTTTCACTGTCGCAACCAAGATTCGCAAAATGGCGT 845  
229 rIleLeuSerPheLeuIleLeuTyrrPheSerHisCysMetIleIysValP 246  
846 TGTTTTAAAGTATTTCAACGGCCTATTTTAAACAGTGTGTCAATGCTGTTA 895  
246 heLeuSerSerLeuIysPheHisIleLeuArg 256  
896 TATATGCTGTGCTAAATCAAAACTTCGCAAG 927  
seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA197.DAT:179065  
seq\_documentation\_block:  
ID: TY9065 standard; cDNA to mRNA; 1312 BP.  
XX



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 11:39:04 ; Search time 65.02 Seconds  
(without alignments)  
82.577 Million cell updates/sec

Title: US-09-510-332-1

Perfect score: 1521  
Sequence: 1 MLESHTLIYFLAVIQFLG.....GNPKLQNAKKFLHSKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/PCRTUS\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/backup1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	8.1	309	1	US-08-118-270-56
2	123.5	8.1	309	4	PCT-US93-08528-56
3	123.5	8.1	348	2	US-08-390-000A-8
4	122.5	8.1	401	2	US-08-820-521-2
5	120.5	7.9	327	3	US-08-748-506-22
6	120.5	7.9	327	3	US-08-748-506-23
7	120.5	7.9	342	3	US-08-785-928-1
8	120.5	7.9	342	3	US-08-728-603-17
9	118.5	7.8	524	2	US-08-928-692-12
10	118.5	7.8	420	2	US-08-466-103A-2
11	117.5	7.7	327	3	US-08-748-506-24
12	113.5	7.5	358	2	US-08-458-970A-11
13	107	7.0	286	4	US-08-118-270-65
14	107	7.0	286	4	PCT-US93-08528-65
15	104.5	6.9	359	4	US-08-148-209A-4
16	103.5	6.8	327	3	US-08-748-506-14
17	103	6.8	342	3	US-08-988-876-9
18	103	6.8	380	3	US-08-153-848-40
19	103	6.8	380	3	US-09-299-843A-40
20	103	6.8	380	4	PCT-US93-11153-40
21	100.5	6.6	316	4	US-08-118-270-44
22	100.5	6.6	316	4	PCT-US93-08528-44
23	100	6.6	277	4	US-08-118-270-68
24	99.5	6.5	277	4	PCT-US93-08528-68
25	99.5	6.5	359	2	US-08-748-485-5
26	99	6.5	325	1	US-08-118-270-51
27	99	6.5	325	1	PCT-US93-08528-51
28	99	6.5	395	4	US-08-097-938-5

29	99	6.5	395	1	US-08-476-000-5	Sequence 5, Appl1
30	99	6.5	395	1	US-08-472-840-5	Sequence 5, Appl1
31	99	6.5	395	2	US-08-476-976-5	Sequence 5, Appl1
32	99	6.5	395	3	US-08-474-410-5	Sequence 5, Appl1
33	99	6.5	399	1	US-08-476-000-61	Sequence 61, Appl1
34	99	6.5	399	1	US-08-472-840-61	Sequence 61, Appl1
35	99	6.5	399	2	US-08-476-976-61	Sequence 61, Appl1
36	99	6.5	399	2	US-08-474-410-61	Sequence 61, Appl1
37	98.5	6.5	372	1	US-08-202-056-5	Sequence 5, Appl1
38	98.5	6.5	372	1	US-08-076-093A-6	Sequence 5, Appl1
39	98.5	6.5	372	1	US-08-701-265-6	Sequence 6, Appl1
40	98.5	6.5	372	2	US-08-284-586-6	Sequence 6, Appl1
41	98.5	6.5	372	2	US-08-805-478-6	Sequence 6, Appl1
42	98.5	6.5	372	2	US-08-802-627A-6	Sequence 6, Appl1
43	98.5	6.5	372	2	US-08-801-238-6	Sequence 6, Appl1
44	98.5	6.5	372	2	US-08-801-228-6	Sequence 6, Appl1
45	98.5	6.5	372	3	US-09-104-296-6	Sequence 6, Appl1

## ALIGNMENTS

```
RESULT 1
US-08-118-270-56
; Sequence 56, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSER: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-118-270-56

Query Match      8.1% Score 123.5, DB 1, Length 309;
Best Local Similarity 21.1% Pred. No. 4.3e-05;
Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps 17;
OY 1 MLESHTLIYFLAVIQFLGIFTNGIIVVNGIDILIRKRM-APLDLLLSCLAVSRIFLQ 59
```

```

Db 2 MAAVM---FLIIVGFPINFLITLVTV-----QHKRLRPIVYIILNLAVADLFMV 50
QY 60 LFIIVVYVIFIEFIFMCSANCAILLFINEL-----ELMLATWLGVPYCAKVASVR 111
Db 51 LGGFTSLIYSLHGYFVFGPTGCLGFFATLGGELALMSLMLAIERVYVYVCKPMSNFRF 110
QY 112 ---HPL-----FIWLMKRISKLV-----WMILGSLVYVSMICVFSKAGFVNYFLR--- 157
Db 111 GENHAIMGVAFVW-MALACAAPPIAGW-----SRY-----IPEGLOQSC 149
QY 158 --KFSQNTAQEDTLAIOISFVAEFSVPLILFPAVLLIFSLGRTRMRNTVAGS 215
Db 150 GIDVYTLKPEVNE---SPVIYMEVVFHTIPLILF-FCYQOLVFTVKEAAAOQES-ATT 204
QY 216 RVPGRGAPISALSLISFLIFESHCMIKVPLSLKPHIRRIPL-----FILLVIGIYPS 271
Db 205 QKAKEVTBMVYIIMVAFILCWVPYASVAFIYTHQGSNFGPIFMRIAPFAKSAIY-- 262
QY 272 GHSLLILGNPKLKNAKKFLHSKCC 298
Db 263 -NPVIYIIFNKOF---RNCMLQLIC 284

```

## RESULT 2

```

PCT-US93-08528-56
; Sequence 56, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NETMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELETYPE: 248633
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-56

```

```

Query Match 8.1%; Score 123.5; DB 4; Length 309;
Best Local Similarity 21.1%; Pred. No. 4.3e-05;
Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps 17;
QY 1 MLESLLIYFLAVIOQLGIFNGIIVVNGIDLKHKRM-APDLLSLCLAVSRIFLQ 59
Db 1 MLESLLIYFLAVIOQLGIFNGIIVVNGIDLKHKRM-APDLLSLCLAVSRIFLQ 59

```

```

Db 2 MAAVM---FLIIVGFPINFLITLVTV-----QHKRLRPIVYIILNLAVADLFMV 50
QY 60 LFIIVVYVIFIEFIFMCSANCAILLFINEL-----ELMLATWLGVPYCAKVASVR 111
Db 51 LGGFTSLIYSLHGYFVFGPTGCLGFFATLGGELALMSLMLAIERVYVYVCKPMSNFRF 110
QY 112 ---HPL-----FIWLMKRISKLV-----WMILGSLVYVSMICVFSKAGFVNYFLR--- 157
Db 111 GENHAIMGVAFVW-MALACAAPPIAGW-----SRY-----IPEGLOQSC 149
QY 158 --KFSQNTAQEDTLAIOISFVAEFSVPLILFPAVLLIFSLGRTRMRNTVAGS 215
Db 150 GIDVYTLKPEVNE---SPVIYMEVVFHTIPLILF-FCYQOLVFTVKEAAAOQES-ATT 204
QY 216 RVPGRGAPISALSLISFLIFESHCMIKVPLSLKPHIRRIPL-----FILLVIGIYPS 271
Db 205 QKAKEVTBMVYIIMVAFILCWVPYASVAFIYTHQGSNFGPIFMRIAPFAKSAIY-- 262
QY 272 GHSLLILGNPKLKNAKKFLHSKCC 298
Db 263 -NPVIYIIFNKOF---RNCMLQLIC 284

```

## RESULT 3

```

US-08-390-000A-8
; Sequence 8, Application US/08390000A
; Patent No. 5985583
; GENERAL INFORMATION:
; APPLICANT: Sealfon, Stuart C.
; TITLE OF INVENTION: Cloning and Expression of
; TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,000A
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELETYPE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-390-000A-8

```

```

Query Match 8.1%; Score 123.5; DB 2; Length 348;
Best Local Similarity 20.8%; Pred. No. 4.9e-05;
Matches 68; Conservative 63; Mismatches 125; Indels 71; Gaps 16;
QY 1 MLESLLIYFLAVIOQLGIFNGIIVVNGIDLKHKRM-APDLLSLCLAVSRIFLQ 59
Db 39 MAAVM---FLIIVGFPINFLITLVTV-----QHKRLRPIVYIILNLAVADLFMV 87

```

QY 60 LFIYVNVIVIFIEIFIMCSANCAILLFI-----NELEW-----LATWLGVCYCAKVASVR 111  
 DB 88 LGGFTSTLYTSLHGIVFPGTCNLEGFATLGGELTALMSLVLAERVYVCKPMSNR 147  
 QY 112 ----HPL----FTW----LKMRLSKLVPMMILGSLLYSMICVFSKYACGMPYFLR--- 157  
 DB 148 FGENHAIMGVAFWVWALMACAPPLAGW-----SRY-----IPEGLOCSG 187  
 QY 158 --KFESONATIOCEDPLAIOIESFVAEFSPPLILFLEAVULLIESLGRHTRORNTVAGS 215  
 DB 188 GIDVYTLKPEVNE---SFYIMFVVFHFTIPMLIFECGQLVFTYKEAAQOQES-ATT 243  
 QY 216 RYDGRGAPISALSLISFLILYFSCMIKIVLSLKFHRRFPL----FELLVIGIYPS 271  
 DB 244 QKAKEKTRAVIIMVIAFLICWVPYASVAFYIFTHOGSNGPIFMTIPAFKSAAY-- 301  
 QY 272 GHSLLILGPNKLNKQNAKFLHSKCC 298  
 DB 302 -NPVIYIMMKQF-----RNCMLTTICC 323

RESULT 4  
 US-08-820-521-2  
 ; Sequence 2, Application US/08820521  
 ; Patent No. 5942416  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bergsma, Derek  
 ; APPLICANT: Ganesh, Sathe  
 ; APPLICANT: Fuelterer, Wendy  
 ; APPLICANT: Mao, Joyce  
 ; TITLE OF INVENTION: CDNA CLONE HNFY20 THAT ENCODES  
 ; TITLE OF INVENTION: A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/820,521  
 ; FILING DATE: 19-MAR-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Han, William T  
 ; REGISTRATION NUMBER: 34,344  
 ; REFERENCE/DOCKET NUMBER: GH50011  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-5219  
 ; TELEFAX: 610-270-4026  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 401 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-820-521-2  
 Query Match 8.1%; Score 122.5; DB 2; Length 401;  
 Best local Similarity 24.0%; Pred. No. 7.3e-05;

Matches 71; Conservative 53; Mismatches 93; Indels 79; Gaps 18;  
 QY 10 FLAVIOLGIFGTNGIIVVNGIDLIKRRMAPLDLISLAVSRIFLOLFIYVAVIY 69  
 DB 73 FSVYLTFELVGLPLN-LTALVVFVKLQRRPVA--VDVYLLMLTASDILLLELFFRVEA 130  
 QY 70 I-----FIEEFIMCSANCAILLFELNELMTWLGVCYCAKVASRRHPLTLMR--- 121  
 DB 131 ANGMHPLFELICPLSG--FIEFTIYL-TALFLAASIERFLSAHPL--WYTRPRLG 185  
 QY 122 ---ISKLVPMILG---SLLYVSMIC--VFHSKYAGFVWPYFLKFSQNTIO---KED 170  
 DB 186 QAGIVSVACMLASAHCSVYVYIEFGDISHSQ-----GTNGCYCLEFRKD 231  
 QY 171 TLAI-----QIESFVAEFSPPLILFLEAVULLIESLGR---HTRORNTVAGSRVGRGAP 223  
 DB 232 QLAILLPLVLENAVVLVPLLTITSYCSRLVWILGRGSHROR----- 277  
 QY 224 ISALL--SLISFLIY---FSH-----C-----MIKVLFLSKFHRRFPLF 261  
 DB 278 VAGLLATILNLFVCGPYNVSHVGYIGCESPAMRIYVTLTLNSCVDPPVYF 333

RESULT 5  
 US-08-748-506-22  
 ; Sequence 22, Application US/08748506  
 ; Patent No. 6159707  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ronnelt et al.  
 ; TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Leydig, Voigt & Mayer, Ltd.  
 ; STREET: Two Prudential Plaza, Suite 4900  
 ; CITY: Chicago  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60601-6780  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/748,506  
 ; FILING DATE: 08-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/033,751  
 ; FILING DATE: 09-NOV-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; REFERENCE/DOCKET NUMBER: 74940  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-616-5700  
 ; TELEFAX: 312-616-5600  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 327 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-748-506-22

Query Match 7.9%; Score 120.5; DB 3; Length 327;  
 Best local Similarity 23.4%; Pred. No. 9.2e-05;  
 Matches 64; Conservative 48; Mismatches 110; Indels 51; Gaps 14;  
 QY 5 HT-IIFYFLAVIOLGIFGTNGIIVVNGIDLIKRRMAPLDLIS-----CLAVSRIF 57  
 DB 28 HRLIFLFLHLAVLASLGMNLITITICVD---HRLQTPMVFILSMSSVCECITTVI 84

QY 58 LQLEFYVNVIVF---FIEFIMCSANCAILLFNELEMLATWLGVF---YCAKVASV 110  
 Db 85 POLL-----TIISGRKQIPFMACISOAFVLYVGATGFEL---LGVLSIDRFALICKPL 136  
 QY 111 RHPLEFMKMRISKLVPMWMLGSLLYVS-MICVPHSKYAG-FMVPYFLRKFFSONATIQK 168  
 Db 137 HYPIMSPRMCFLLVTVSLVGLFPMASPVYMLQSOFYCGNIIIPHFCD-FGPLANLSC 195  
 QY 169 EDTLAIQIFSVFAEFVSVPLLEFAVALL-IFSLGHRTRQMRNTVAGSRVPGCAPISAL 227  
 Db 196 SETSIEMLEFTEL-----AIVLFAASLILAFAYS-----NIVVTIVRLPSARERORA- 243  
 QY 228 LSLSLFLLYFHSCHMIKVFSLSLKFIHRRIFEL 260  
 Db 244 -----FSTCSSHLIYLSIMYGSACAFIYL 266

RESULT 6  
 US-08-748-506-23  
 ; Sequence 23, Application US/08748506  
 ; Patent No. 6159707  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ronnett et al.  
 ; TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
 ; STREET: Two Prudential Plaza, Suite 4900  
 ; CITY: Chicago  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60601-6780  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/748,506  
 ; FILING DATE: 08-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/033,751  
 ; FILING DATE: 09-NOV-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; REFERENCE/DOCKET NUMBER: 74940  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-616-5600  
 ; TELEFAX: 312-616-5700  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 327 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-748-506-23

Query Match 7.9%; Score 120.5; DB 3; Length 327;  
 Best Local Similarity 23.4%; Pred. No. 9,2e-05;  
 Matches 64; Conservative 48; Mismatches 110; Indels 51; Gaps 14;  
 QY 5 HLIIYFLAVIOFLGFTNGCIIVVNGIDILKHKRMAPLDDLIS-----CLAVSRIF 57  
 Db 28 HURILFELLHLAYLASLGMNLIITITVD---HRLQIPMFFLISMSSVECCFTTIVT 84  
 QY 58 LQLEFYVNVIVF---FIEFIMCSANCAILLFNELEMLATWLGVF---YCAKVASV 110  
 Db 85 POLL-----TIISGRKQIPFMACISOAFVLYVGATGFEL---LGVLSIDRFALICKPL 136  
 QY 111 RHPLEFMKMRISKLVPMWMLGSLLYVS-MICVPHSKYAG-FMVPYFLRKFFSONATIQK 168

Db 137 HYPIMSPRMCFLLVTVSLVGLFPMASPVYMLQSOFYCGNIIIPHFCD-FGPLANLSC 195  
 QY 169 EDTLAIQIFSVFAEFVSVPLLEFAVALL-IFSLGHRTRQMRNTVAGSRVPGCAPISAL 227  
 Db 196 SETSIEMLEFTEL-----AIVLFAASLILAFAYS-----NIVVTIVRLPSARERORA- 243  
 QY 228 LSLSLFLLYFHSCHMIKVFSLSLKFIHRRIFEL 260  
 Db 244 -----FSTCSSHLIYLSIMYGSACAFIYL 266

RESULT 7  
 US-08-785-928-1  
 ; Sequence 1, Application US/08785928  
 ; Patent No. 6087115  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gershengorn, Marvin C.  
 ; APPLICANT: Arvanitakis, Ieandros  
 ; APPLICANT: Geras-Raaka, Elizabeth  
 ; APPLICANT: Cesalman, Ethel  
 ; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED  
 ; TITLE OF INVENTION: RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/785,928  
 ; FILING DATE: 22-JAN-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Goldman, Michael L.  
 ; REGISTRATION NUMBER: 30,727  
 ; REFERENCE/DOCKET NUMBER: 19603/1320  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (716) 263-1304  
 ; TELEFAX: (716) 263-1600  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 342 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-785-928-1

Query Match 7.9%; Score 120.5; DB 3; Length 342;  
 Best Local Similarity 22.4%; Pred. No. 9,7e-05;  
 Matches 59; Conservative 56; Mismatches 95; Indels 53; Gaps 14;  
 QY 11 LVAIYPLFLGFTNGCIIVVNGIDILKHKRMA-PLDLLSLAVSRIFLQLEFYVNVIV 69  
 Db 53 ILSLI-FLIIVNGGLVYI---FCKHRSRAGADIDILGICLNSICLSI-LAEVL 105  
 QY 70 IFETEFIMCSANCAILLFNELEMLATWLGVFCAVAVSRHPLFI-----MLKMRISK 124  
 Db 106 MELPNIISTGCLGLEIF-----YLVYVIDIFSVCVSLVRYLLAVYSTRSNPK---KQ 158  
 QY 125 LVPMMILGSLIYVMT-----CVFHSKYAGFMVPYFLRKFFSONATIQKED-TLAIQIFS 178  
 Db 159 SLGAVLTSAAULLIALVLVSGDACRHRSHVVD---PVSKQAMCYENAGNMTADMRLHVRTVS 215

OY 179 FVAEFSVPLILFLFAVLLIFSLG-----RHR-OMRNTVAGSRVPGAPISALLSLISF 233  
 DB 216 VTAGEFLPL-----ALLILFYALTWCVRRTKIOARKRVG-----YIVAV 256  
 OY 234 LILYFSHQMIRKVLSSLKFIHR 256  
 DB 257 VILFFVFCPPYHVLNLDLTLRR 279

RESULT 8  
 US-08-728-603-17  
 ; Sequence 17, Application US/08728603  
 ; Patent No. 6093806  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Caesarman, Ethel  
 ; APPLICANT: Knowles, Daniel M.  
 ; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED  
 ; TITLE OF INVENTION: HERPESVIRUS  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/728,603  
 ; FILING DATE: 10-OCT-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BRAMAN, SUSAN J.  
 ; REGISTRATION NUMBER: 34,103  
 ; REFERENCE/DOCKET NUMBER: 19603/720  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 716-263-1636  
 ; TELEFAX: 716-263-1600  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 342 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-728-603-17

Query Match 7.9%; Score 120.5; DB 3; Length 342;  
 Best Local Similarity 22.4%; Pred. No. 9.7e-05;  
 Matches 59; Conservative 56; Mismatches 95; Indels 53; Gaps 14;

OY 11 LLAIVQFLGIFNGIIVVNGIDLKHKRMA-PLDLLSLCLAVSRIFLOLFYFVNVIV 69  
 DB 53 ILSDI-FLINVLNGLVITYI-----FCKHRSRAGADILLGLICLSICUSISL--LAEVL 105  
 OY 70 IFLEFIMCSANCAILFIFINELMLATWLGVEYCAKVASVRHPLF-----WLKMRISK 124  
 DB 106 WLEFPIIISIGLCLEIF-----YILVYVIDIFSVCVSLVRYLLAVAYSRSPK---KQ 158  
 OY 125 LVPMMIGSLIYVSM-----CVFHSKYAGPMYPIFLRKFSQNAITQKED-TLAQIIFS 178  
 DB 159 SLGWVLTSAALLIALIVSGDGRHSRVVD---PVSQAMCYENAGNMTADMRLHVTVS 215  
 OY 179 FVAEFSVPLILFLFAVLLIFSLG-----RHR-OMRNTVAGSRVPGAPISALLSLISF 233  
 DB 216 VTAGEFLPL-----ALLILFYALTWCVRRTKIOARKRVG-----YIVAV 256  
 OY 234 LILYFSHQMIRKVLSSLKFIHR 256

DB 257 VILFFVFCPPYHVLNLDLTLRR 279

RESULT 9  
 US-08-928-692-12  
 ; Sequence 12, Application US/08928692  
 ; Patent No. 5958727  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brody, Howard  
 ; APPLICANT: Yaver, Deborah S.  
 ; APPLICANT: Lamsa, Michael  
 ; APPLICANT: Hansen, Kim  
 ; TITLE OF INVENTION: Methods for Modifying the Production of  
 ; TITLE OF INVENTION: a Polypeptide  
 ; NUMBER OF SEQUENCES: 80  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 59587270 No. 5958727disk of No. 5958727th America, Inc.  
 ; STREET: 405 Lexington Avenue  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10174  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/928,692  
 ; FILING DATE: 12-SEPT-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lambiris, Elias J.  
 ; REGISTRATION NUMBER: 33,728  
 ; REFERENCE/DOCKET NUMBER: 4944,200-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 524 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 5958727e  
 ; US-08-928-692-12

Query Match 7.8%; Score 118.5; DB 2; Length 524;  
 Best Local Similarity 24.8%; Pred. No. 0.00026;  
 Matches 55; Conservative 30; Mismatches 64; Indels 73; Gaps 11;

OY 6 LIIYVLAIVQFLGIFNGI-----IVVYNGIDLKHKRMADLLSLCLAVSRIFLO 59  
 DB 304 ILVALMLHVAOOFST--NGIFITYSTISIFOTAGIS-----KRYVNTIGGAVNMYETA 354  
 OY 60 LFIFVYVIVIFLEFIMCSANCAILFIFINELMLATWLGVEYCAKVASVRHPLFTWK 119  
 DB 355 VSEFLVER-----AGRRSLFLIOMSGMFVCA-----IFMSVG 386  
 OY 120 KRISLVPMMIGSLIYVSMICVFHSKYAGM-----VPEFL-RKFSQNAITQKEDTL 172  
 DB 387 LVLLNKFSWM-----SYVSMIAIF--LFVSFPEIGPGFIEMFVAEFSSQ---PRPAL 436  
 OY 173 AIQIIFS-----FVAEFSVPLILFLFAVLLIFSL 201  
 DB 437 AIAAFSNTMCNTVALCPQYINDPGCPYVFFLFAVLLAFTL 478

RESULT 10  
 US-08-466-103A-2  
 ; Sequence 2, Application US/08466103A





```

; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF INVENTION: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 286 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-65

```

```

Query Match      7.0%; Score 107; DB 4; Length 286;
Best Local Similarity 21.6%; Pred. No. 0.0019;
Matches 66; Conservative 52; Mismatches 131; Indels 56; Gaps 11;

QY 8 IYFLAVIOELGFTNGITIVVNGIDILKHKRMARLDLLSCLAVSRIFQLFITYVNV 67
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 LFLFLSLXVVLVTEMLTIIRNHPFLHKPM-----YFLFLLEIYVTV 47

QY 68 IVYFEFFIMCSANCAITL--FTNELEMLATWLGVEYCAKVA-----SVRHPLF-- 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 48 TIRPLMGFISKEHNGQLISFFACMTQLYFFELGCTECVLLAVMAYDRVVAICHLHP 107

QY 116 IMLKMRISKLVPMWILGSLVYVMICVFNHKGAFNVPFLRKFFSQNATIQEDFLAIQ 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 108 VIYSSRIZVIGSM---AGRGISMVAVFLISRLSYCGPNTINHFC-----DIVSPLL 156

QY 176 TFSFVAEFSVPLLIFFAVLLIFSLGRHTRQMRNTVAGS---RVP---GRGAPISALLS 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 157 NLSCTDMSTAEIADDFVLAIFILGPL-----SVTASVYKRISSAGRKRAESTCAS 207

QY 230 ILSFLIYFESHCHIKVFLSKAFHTRFT--FLFPLVIGIYPSGSHLLILGNPKLON 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 208 HLTVVLIIFVA---ASIFIVARPKALSAFTDNKLVAVLAVIYVLEFNPIIYCLRNQVKRA 264

QY 288 AKRFL 292
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 265 LRRTL 269

```

```

RESULT 15
US-08-148-209A-4
; Sequence 4, Application US/08148209A
; Patent No. 5556780
; GENERAL INFORMATION:

```

```

; APPLICANT: Dzau, Victor J
; APPLICANT: Mukoyama, Masashi
; TITLE OF INVENTION: TYPE-2 ANGIOTENSIN II RECEPTOR AND GENE
; NUMBER OF INVENTION: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,209A
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58491-1/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 396-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-148-209A-4

```

```

Query Match      6.9%; Score 104.5; DB 1; Length 359;
Best Local Similarity 21.3%; Pred. No. 0.0044;
Matches 70; Conservative 58; Mismatches 118; Indels 83; Gaps 17;

QY 5 HLIIYFLAV---IOFLGIFTNGITIVVNGIDILKHKRMARLDLLSCLAVSRIFQLF 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 24 HSYIFEMIPTLSIIFVGVIFGNSLVVIYTF-YMKLTVASVFLNLALADLCFLTLTP 82

QY 62 IYVNVVIVFEIEF--IMCSANCAITLFTNELEMLATWLGVEYCAKVASRHP----- 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 83 LMAVTTAMETIRMPFENHLCIKLASASVSFNLYASVFLTLCLSI---DRYLAVHPKSRRL 139

QY 114 -----LFTWLKMRISKLVPMWILGSLVY---MICEFH----- 144
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 140 RTMLVAKVTCIITIMMAGIASL-PAVIRHNVYFIENNTITCAFIYESRNLTLPIGLT 198

QY 145 SKYAGFNVPYF-----LAKFFSQNATIQED---TLAIOESFAVEFS-VP 186
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 199 KNILDFLEPFLIILTSYLIMKALKAKAYEIQKNRPNDIPRIIMAYLFFP---FSWVP 255

QY 187 LIIFLAVLLIFSLG-RHTRQMRNTVAGSRVPGAPISALLSILSFLIYFESHCHIKV 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 256 HQIFTF-LDVLQIGVHDKADI-IDVTAMP-----ITICIAVFNNCLNPL 300

QY 246 FLSSLKPHIRRFILFPLVIGIYPSGHS 274
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 FYGFLGKKKKRY-FLQLLKVIYIPRAKSHS 328

```

```

Search completed: March 15, 2001, 12:57:20
Job time: 4696 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 15, 2001, 12:32:39 ; Search time 76.8 Seconds  
(without alignments)  
133.124 Million cell updates/sec

Title: US-09-510-332-1

Perfect score: 1521  
Sequence: 1 MESHLLIIFLLAVIQFLG.....GNPKIKAKAKFLHKKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_36.\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
22: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263.5	17.3	256	Y42388	Amino acid sequence
2	124.5	8.2	348	R38483	Rhodopsin protein.
3	124.5	8.2	348	R38483	Rhodopsin. Homo s
4	123.5	8.1	309	R48735	G-protein coupled
5	123.5	8.1	309	R48735	G-protein coupled
6	122.5	8.1	401	W59924	Human 7-transmembr
7	120.5	7.9	342	W37976	Kaposi's sarcoma a
8	118	7.8	354	W25926	Xenopus melatonin
9	118	7.8	354	W25926	Xenopus melatonin
10	117.5	7.8	420	R88409	High-affinity mela
11	117.5	7.7	310	W19105	Rat pheromone rece
12	117.5	7.7	522	W17836	Rat glucose transp

13	117.5	7.7	690	Y13352	Amino acid sequenc
14	117.5	7.7	690	Y13352	PRO288 Polypeptide
15	116	7.6	349	P90554	Bovine rhodopsin.
16	114	7.5	123	Y27593	Human secreted pro
17	112.5	7.4	354	Y57086	Rhodopsin amino ac
18	112.5	7.4	476	Y70157	Staphylococcus aur
19	109.5	7.2	362	Y16630	Human putative Adr
20	109.5	7.2	389	Y19909	B. burgdorferi ant
21	109.5	7.2	449	Y19908	B. burgdorferi ant
22	108	7.1	337	Y19563	Human G protein co
23	107.5	7.1	312	R27875	Olfactory receptor c
24	107	7.0	286	R48744	G-protein coupled
25	107	7.0	286	W02716	G-protein coupled
26	106.5	7.0	372	W86323	Kidney injury asso
27	104	6.8	342	R25834	Guinea pig platele
28	103.5	6.8	327	W1666	Rat spermatozoa
29	103.5	6.8	437	Y34786	Amino acid sequenc
30	103.5	6.8	690	Y27593	Human G protein co
31	103	6.8	337	W85047	Rat pheromone rece
32	103	6.8	337	W85047	A human 7-transmem
33	103	6.8	337	W75799	Human 7-transmembr
34	103	6.8	342	W12243	Guinea pig platele
35	103	6.8	380	R53750	Seven transmembran
36	103	6.8	380	W48731	Human R20 seven tr
37	103	6.8	380	Y24952	G-protein-coupled
38	103	6.8	380	Y88442	G-protein-coupled
39	103	6.8	380	Y87466	Human APJ receptor
40	102	6.7	780	Y41089	Human lectomedin-1
41	102	6.7	1021	Y77728	Human G protein-co
42	102	6.7	1114	Y41092	Peptide Seq ID No:
43	102	6.7	1123	Y41086	Human lectomedin-1
44	102	6.7	1177	Y41085	Human lectomedin-1
45	102	6.7	1403	Y41087	Human lectomedin-1

#### ALIGNMENTS

RESULT	1	ALIGNMENTS
Y42388	Y42388 standard. Protein: 256 AA.	
XX	Y42388;	
AC	09-DEC-1999 (first entry)	
XX		
DE	Amino acid sequence of p127.1.	
XX		
KW	secreted protein; cDNA library; clone; transmembrane protein;	
KW	signal sequence cloning; hybridization cloning; gene therapy;	
KW	receptor.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	8..20
FT	Protein	/label= leader/signal peptide
FT		21..256
FT		/label= Mature protein
FT	Misc-difference	64
FT		/label= unknown
FT		/note= "encoded by twa"
XX		
XX	W09942470-A1.	
PN		
XX	26-AUG-1999.	
PD		
XX	18-FEB-1999;	99WO-US03458.
PF		
XX	18-FEB-1998;	98US-0075038.
PR	17-FEB-1999;	99US-0251600.
XX		
PA	(GENY ) GENETICS INST INC.	

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Werberg D;  
PI Treacy M, Agostino MJ, Steinger RJ;  
XX WPI; 1999-518580/43.  
DR N-PSDB; 220861.  
XX  
PT New polynucleotides encoding human secreted proteins used for  
PT therapeutic, diagnostic and research purposes.  
XX  
PS Claim 27; Page 114-115; 125pp; English.

XX This is the amino acid sequence of the p127-1 protein, which is  
CC derived from the p127-1 clone isolated from a human adult blood  
CC (lymphoblastic leukemia MOLT-4) cDNA library.  
CC The pns and proteins of the invention are predicted to have biological  
CC activities which would make them suitable for treating, preventing or  
CC ameliorating medical conditions in humans and animals, although no  
CC supporting data is given. Suggested activities include nutritional  
CC activity, cytokine and cell proliferation/differentiation activity,  
CC immune stimulating (e.g. as vaccines) or suppressing activity,  
CC hematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumor invasion suppressor activity, and tumor  
CC inhibition activity. The pns are also stated to be useful for gene  
CC therapy. Other activities include inhibiting the growth, infection or  
CC function of bacteria, fungi, viruses and other parasites; effecting  
CC bodily characteristics such as, e.g. weight, color, skin, etc.;  
CC effecting biorhythms or cardiac cycles; enhancing fertility; treatment  
CC of depression; treatment of pain; hormonal or endocrine activity.  
XX  
SQ Sequence 256 AA;

Query Match 17.3%; Score 263.5; DB 20; Length 256;  
Best Local Similarity 28.8%; Pred. No. 7.9e-23;  
Matches 74; Conservative 58; Mismatches 74; Indels 51; Gaps 8;

QY 7 IIVFLAVIOFLGIFTNGITVVNGIDLIKRRKMAPDLILSLCAVSRIFQLFFYYN 66  
DB 7 IISLLVVFVIGNFANGFIVLNSLEWKQKISFADQLTALAVALRYGL-----LW 60  
QY 67 VIVVFIEFIMCSANCAILFT-----NELEMLATWLCVFCYCAKAVSRHPLFTW 117  
DB 61 VILXHYAEVLNPGSYSLGVRLCTLNAAVCTNFSIWAETSLIFYLILKIANFSNFIHL 120  
QY 118 LKMRISKLVPMWILGSLLYSMIC-----VFHSKAGFNVVYFLRKFSQATIQ 167  
DB 121 LKRIKAVPVLIIIGLILF--LVCHLVVVMDESMTWCKEYEGNV-----SWEIKIS 169  
QY 168 KEDTLAIQIFSEVAEFSVPLILFLFAVLLIFSLGRHTRQMRNTVAGSRVPGGAP--- 223  
DB 170 dphldemvctlanl-1pftclslslflllclclckhkkmm-----qfhgyspsdnt 221  
QY 224 ---ISALSTISFLICY 237  
DB 222 kvhkalqvtcsfllilf 238

RESULT 2  
R38483  
ID R38483 standard; Protein: 348 AA.  
XX R38483;  
XX  
XX 11-NOV-1993 (first entry)  
XX  
XX Rhodopsin protein.  
XX  
XX Human; rhodopsin; mutant; retinal degeneration; primer; probe;  
XX hereditary.  
XX

OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 249  
FT /note= "Sequence unclear at this position"  
XX

PN MO9312134-A.  
PD 24-JUN-1993.  
XX  
XX 08-DEC-1992; 92WC-US10536.  
XX  
XX 11-DEC-1991; 91US-0805123.  
XX

PA (HARD) HARVARD COLLEGE.  
PI Berson EL, Dryja TP;  
XX  
XX WPI; 1993-214088/26.  
DR N-PSDB; Q43543.

PT Probe or primer contg. sequence of human retinal degeneration  
PT slow protein mutant - used to diagnose hereditary retinal  
PT degenerative diseases  
XX  
PS Disclosure: Fig 1; 56pp; English.

CC The sequence given represents the human rhodopsin protein. Mutant  
CC rhodopsin proteins cause retinal degeneration. The DNA encoding the  
CC mutant DNA sequences may be identified using primers/probes described  
CC in the invention (see also Q43545-48) and may be used to diagnose  
CC hereditary retinal degeneration.  
XX

SQ Sequence 348 AA;

Query Match 8.2%; Score 124.5; DB 14; Length 348;  
Best Local Similarity 20.8%; Pred. No. 2.1e-06;  
Matches 68; Conservative 63; Mismatches 125; Indels 71; Gaps 16;

QY 1 MESHIIIFLLAVIOFLGIFTNGITVVNGIDLIKRRKMAPDLILSLCAVSRIFQL 59  
DB 39 mlaaym--fllivlgfpinfillyvtv-----qnhkrlrpinylilnlavadihm 87  
QY 60 LTFYVNVIVIFIEFIMCSANCAILFT-----NELEMLATWLCVFCYCAKAVSR 111  
DB 88 lggfsclytslshyfvfgptcnglegfctlggelalvslvdleryvvvckpmsnfr 147  
QY 112 ---HPL---FIW---LKMRIKLVPMWILGSLLYSMICVFSKAGFNVVYFLR--- 157  
DB 148 fgenhalingvafitwmalacaaplagw-----ary-----lpegllqsc 187  
QY 158 --KFSQNAITQKEDTLAIQIFSEVAEFSVPLILFLFAVLLIFSLGRHTRQMRNTVAGS 215  
DB 168 gldyrlkpevne---slvlymtvvhftlpmllifcyygqlvtvkaaaagqges-alc 243  
QY 216 RVPGRGAPISALLSILSLILYFESHCMIKVFLSLKFIHRIFL-----FFITVGIYPS 271  
DB 244 gkaekxvtlrmvliawlaflilcwpyasvafylfthgnsnfgplfnclpafefkasaaly-- 301  
QY 272 GHSLLILGKPKLAKONAKKFLHSHKCC 298  
DB 302 -npvilylmmnxqf---trncmltllic 323

RESULT 3  
R93116  
ID R93116 standard; Protein: 348 AA.  
XX R93116;  
XX  
XX 06-JUL-1996 (first entry)  
XX  
XX

[illegible]

```

Db 244 qkakevtrlmvlnvialfllcwpyasvafifchgsnfgplmtfipafafxaaay-- 301
OY 272 GHSLILIGNPKLONAKKFLHSKCO 299
: : | : | : : : | : | :
Db 302 -npviylmmkqf----rncmltliccr 324

RESULT 4
R48735
ID R48735 standard; Protein; 309 AA.
XX
XX R48735;
AC
XX
XX 07-JUN-1996 (first entry)
DE
XX G-protein coupled human rhodopsin protein.
XX
XX
XX G-protein coupled receptor; ligand binding assay; transmembrane domain;
XX psychotic disorder; schizophrenia; dopamine; CAMP; adenosine; thrombin;
XX muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
XX rhodopsin; opsin; odorant; cytomegalovirus.
XX
XX Homo sapiens.
XX
XX W09405695-A1.
XX
XX 17-MAR-1994.
XX
XX 09-SEP-1993; 93WO-US08528.
XX
XX 10-SEP-1992; 92US-0943236.
XX
XX (UYNY ) UNIV NEW YORK STATE.
XX
XX Murphy RB, Schuster DI;
XX
XX WPI: 1994-101120/12.
XX
XX PT Polypeptides of G-coupled receptor proteins (GPRs) - useful for
XX binding GPR ligands or modulating GPR binding
XX
XX Disclosure; Page 111-112; 160pp; English.
XX
XX
XX Proteins R48665-R48758 represent a range of G-protein coupled receptor
XX proteins selected from CAMP, adenosine, muscarinic acetylcholine, opsin,
XX adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin,
XX odorant, cytomegaloviral and other G-protein coupled receptors. The
XX receptor proteins were used to design polypeptides, pref. based on the
XX transmembrane domains, for use in G-protein coupled receptor ligand
XX binding assays. The polypeptide fragments retain biological activity
XX such as binding a GPR ligand or modulating GPR ligand binding to a GPR
XX (see R48755-R48758, R50569-R50807 and R89189-R89195 for examples of
XX polypeptide fragments). The polypeptide fragments can be used in
XX compositions for treating subjects suffering from a pathology related
XX to a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
XX
XX
XX Sequence 309 AA:

Query Match 8.1%; Score 123.5; DB 15; Length 309;
Best Local Similarity 21.1%; Pred.No.2.3e-06;
Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps 17;

OY 1 MESHILITVELLAVIOFLGIFNGIIVVNGIDLKHKRM-APDILLSCLAVSRIFFQ 59
: : : : | : | : | : | : : : : : : : : : : : : : : : : : : : :
Db 2 mlaaam-----lllvialfipnfllytv-----ghkklltrpnylllnlavaldmfv 50

OY 60 LFIIFYNVVIVFIEFIEFIMCSANCALLFINEL-----ELWLATWLGVFYCAKVASVR- 111
: : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 51 lggftstllyslngyfyfpgtgcnlgeffeatlgelawslvalateryvvvckpmsnfrf 110
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 112 ----HPL-----FIMLKMRISKLV-----WMILGSLVLVSMICVHSKAGFMVYFPR--- 157
: : : : | : | : | : : | : | : : : : : : : : : : : : : : : : :

```

Db 111 genhalmgvaftw-malacaaplagw-----sry-----lpegqsc 149

Oy 158 --KFEQGNATIQEDLAIQIFSVAFSPVLLIFFAVLLIFSLGRTRQNRNTVAGS 215

Db 150 gldyyltkpevne--sflymfvvhtlplllf-fcygqlvftvkeaaqqes-att 204

Oy 216 RVPGRGAPISALSLILSFLLYFSHCMIKVFSLKFIHRIFFL----FFILVIGIYPS 271

Db 205 qkaekvtrmvlmvaflclcwpyasvafylftngsnfgplfmrlpafaksaaly-- 262

Oy - 272 GHSLLILGNPKLQKONAKKFLHSKCC 298

Db 263 -npvyliflnkgf-----rncmlqlcc 284

RESULT 5

W02707

ID W02707 standard; peptide: 309 AA.

XX

AC W02707;

XX

DT 13-NOV-1996 (first entry)

XX

DE G-protein coupled human rhodopsin.

XX

KW G-protein coupled receptor; ligand binding assay; transmembrane domain;

KM schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;

KM muscarinic acetylcholine; endochelin; bombesin; endocrine; rhodopsin;

KM odorant; cytomagalovirus; serotoninergic.

XX

OS Homo sapiens.

XX

PN US5508384-A.

XX

PD 16-APR-1996.

XX

PE 10-SEP-1992; 92US-0943236.

XX

PR 09-SEP-1993; 93US-0118270.

XX

PR 10-SEP-1992; 92US-0943236.

XX

PA (UVMY ) UNTV NEW YORK STATE.

XX

PI Murphy RB, Schuster DI;

XX

DR WPI: 1996-208785/21.

XX

PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.

PT for treating schizophrenia

XX

PS Disclosure: Column 139-140; 184pp; English.

XX

CC Proteins W02657-W02730 represent a range of G-protein coupled receptor

CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,

CC adrenergic, thrombin, endochelin, bombesin, endocrine, rhodopsin, opsin,

CC odorant, cytomagaloviral and other GPR proteins. The receptor proteins

CC were used to design polypeptides, pref. based on the transmembrane

CC domains, for use in G-protein coupled receptor ligand binding assays.

CC The polypeptide fragments retain biological activity such as binding a

CC GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999

CC for examples of polypeptide fragments). The polypeptide fragments can

CC be used in compositions for treating subjects suffering from a pathology

CC related to a GPR abnormality e.g. a psychotic disorder such as

CC schizophrenia.

XX

Sequence 309 AA;

Query Match 8.1%; Score 123.5; DB 17; Length 309;

Best Local Similarity 21.1%; Pred. No. 2.3e-06;

Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps 17;

Oy 1 MLESLLIYFLAVIQFLGIFTNCTIYVVGIDILIRKM-APDLLLSCLAVSRIFLQ 59

Db 2 mlaaym--flilvlgfplnfllyvtv-----ghkkrlrplyllnlnvadalfmw 50

Oy 60 LFIIFYVNVVIFPIEFIMCSANCAILLFTINEL-----ELMLATWLGVEYCAKVASVR- 111

Db 51 lsgftctlytslmgvfvfbptcgnllegffatlggelaalslwaleryvvvckpmnftf 110

Oy 112 ---HPL----FIWLKMRISKLPV---WMILGSLLYVSMICVFHSKYAGFMVPEFLR- 157

Db 111 genhalmgvaftw-malacaaplagw-----sry-----lpegqsc 149

Oy 158 --KFEQGNATIQEDLAIQIFSVAFSPVLLIFFAVLLIFSLGRTRQNRNTVAGS 215

Db 150 gldyyltkpevne--sflymfvvhtlplllf-fcygqlvftvkeaaqqes-att 204

Oy 216 RVPGRGAPISALSLILSFLLYFSHCMIKVFSLKFIHRIFFL----FFILVIGIYPS 271

Db 205 qkaekvtrmvlmvaflclcwpyasvafylftngsnfgplfmrlpafaksaaly-- 262

Oy 272 GHSLLILGNPKLQKONAKKFLHSKCC 298

Db 263 -npvyliflnkgf-----rncmlqlcc 284

RESULT 6

W59924

ID W59924 standard; protein: 401 AA.

XX

AC W59924;

XX

DT 07-DEC-1998 (first entry)

XX

DE Human 7-transmembrane receptor HNFEDY20.

XX

KW HNFEDY20; G-protein coupled receptor; human; infection; HIV; pain;

KM cancer; anorexia; asthma; Parkinson's disease; acute heart failure;

KM hypotension; hypertension; urinary retention; osteoporosis;

KM angina pectoris; myocardial infarction; ulcer; allergy;

KM benign prostatic hypertrophy; psychosis; anxiety; schizophrenia;

KM manic depression; delirium; dementia; mental retardation;

KM dyskinesia; Huntington's disease; Gilles de la Tourette's syndrome;

KM therapy; diagnosis.

XX

OS Homo sapiens.

XX

PN EP666126-A1.

XX

PD 23-SEP-1998.

XX

PE 16-FEB-1998; 98EP-0301122.

XX

PR 19-MAR-1997; 97US-0820521.

XX

PA (SMIK ) SMITHKLINE BEECHAM CORP.

XX

PI Bergsma DJ, Fuechterer WS, Mao JY, Sathe GM;

XX

DR WPI: 1998-482962/42.

XX

PT N-PSDB: V53631.

XX

PT New polynucleotides and polypeptides encoding a novel human

PT 7-transmembrane receptor - useful for diagnosing and treating e.g.

PT cancer, osteoporosis and Parkinson's disease and infections caused

PT by HIV-1 or -2.

XX

PS Claim 1; Page 18-19; 24pp; English.

XX

CC This polypeptide comprises HNFEDY20, a novel human 7-transmembrane

CC G-protein coupled receptor that shows about 30.8% identity in 299

CC amino acid residues with the thrombin receptor. Its amino acid

CC sequence was deduced from an isolated HNFEDY20 polynucleotide

CC sequence (see V53631). The invention relates to HNFEDY20

CC polypeptides and recombinant materials and methods for their

```
DR      MPI: 1998-261008/23.
XX      N-PSDB; V31718 .
PT      Isolated Kaposi's sarcoma-associated herpesvirus proteins -
PT      comprising antigenic membrane protein, G protein coupled receptor
PT      and cyclin protein used to develop products for diagnosis and
XX      therapy
XX      Example 1: Page 51-53; 68pp; English.
XX      The sequence is that of a G-protein-coupled receptor encoded
CC      by ORF74 of Kaposi's sarcoma herpesvirus (KSHV). It can be
CC      used in the diagnosis and treatment of KHSV infection.
XX      Sequence    342 AA:
S0
```

	Query Match	Similarity	Score	DB	Length
Best Local	59	Conservative	7.98	120.56	342
Matches	59	Conservative	7.98	Pred. No. 6e-06	DB 19
				Mismatches	95
				Indels	53
				Gaps	14

  

	Query	Match	Similarity	Score	DB	Length
Best Local	59	Conservative	7.98	120.56	342	
Matches	59	Conservative	7.98	Pred. No. 6e-06	DB 19	
				Mismatches	95	
				Indels	53	
				Gaps	14	

  

	Query	Match	Similarity	Score	DB	Length
Best Local	59	Conservative	7.98	120.56	342	
Matches	59	Conservative	7.98	Pred. No. 6e-06	DB 19	
				Mismatches	95	
				Indels	53	
				Gaps	14	

[illegible]

RESULT	8
ID	W25926 standard; Protein; 354 AA.
AC	W25926;
DT	10-NOV-1997 (first entry)
DE	Xenopus melatonin receptor MEL-1a.
KW	Allele; Xenopus laevis; melatonin; receptor; untranslated region; PCR;
KW	mRNA; half-life; skin; amplification; primer; polymerase chain reaction;
KW	transmembrane domain; cellular signaling; inhibition; adenylyl cyclase;
KW	modulation; intracellular; cyclic GMP; inhibitor; phosphodiesterase.
OS	Xenopus laevis.
FT	
FT	Key
FT	Location/Qualifiers
FT	1..32
FT	/note= "Extracellular N-terminal domain"
FT	33..56
FT	/note= "Transmembrane domain 1"
FT	57..68
FT	/note= "Intracellular domain 1"
FT	69..92
FT	/note= "Transmembrane domain 2"
FT	93..107
FT	/note= "extracellular domain 2"
FT	108..125
FT	/note= "transmembrane domain 3"
FT	126..151
FT	Domain

```

FT FT /note= "intracellular domain 2"  

FT Domain 152..171  

FT FT /note= "transmembrane domain 4"  

FT Domain 172..193  

FT FT /note= "extracellular domain 3"  

FT Domain 194..219  

FT FT /note= "transmembrane domain 5"  

FT Domain 220..243  

FT FT /note= "intracellular domain 3"  

FT Domain 244..268  

FT FT /note= "transmembrane domain 6"  

FT Domain 269..280  

FT FT /note= "extracellular domain 4"  

FT Domain 281..301  

FT FT /note= "transmembrane domain 7"  

FT Domain 302..354  

FT FT /note= "intracellular C-terminal domain"  

  

PN W09704.094-A1.  

XX  

XX PD 06-FEB-1997.  

XX PO  

XX PR 24-JUL-1996; 96MO-FR01167.  

XX PE  

XX PR 24-JUL-1995; 95FR-0008947.  

XX PA (ADIR ) ADIR & CIE.  

XX PI  

XX XX Jockers R, Marullo S, Strosberg AD;  

DR WPI: 1997-122635/12.  

DR N-PSDB: T79063-4.  

XX  

XX New nucleic acid encoding functional melatonin receptor of Xenopus -  

PT for screening for potential (ant)agonists useful for e.g. treating  

PT cardiovascular disease and cancer  

XX  

XX Claim 5; Page 25-26; 62pp; French.  

XX  

XX Sequences W25926-7 represent novel allelic Xenopus laevis melatonin  

CC receptor MEL-1A proteins. This sequence is a new form of the melatonin  

CC receptor designated MEL-1Aa, also known as Mel 1c(alpha). The protein  

CC differs from previously known receptors by being 65 amino acids shorter  

CC and also having 2 different C-terminal amino acids. The protein is  

CC encoded by 2 different alleles (T79063-4) which differ in the 3'  

CC untranslated region (both sequences encode the same protein), which is  

CC thought to affect the half-life of the mRNA. The coding sequence was  

CC isolated from cDNA derived from Xenopus skin RNA and amplified using the  

CC primers T79067-76. The nucleotide sequence encodes a 7 transmembrane  

CC receptor involved in cellular signalling. MEL-1Aa has been shown to  

CC inhibit adenyllyl cyclase, but both proteins can modulate intracellular  

CC cGMP, esp. inhibiting its accumulation induced by an inhibitor of  

CC phosphodiesterase.  

XX  

XX Sequence 354 AA:  

  

OY Query Match 7.8%; Score 118; DB 18; Length 354;  

Db Best Local Similarity 19.4%; Pred. NO. 1,2e-05;  

Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12  

  

OY 2 LESHLLIVFLAVIQGLFINGIIVVNGDILKHKMKP-LDLLSLAVSRIFLQ 60  

| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  

Db 31 ltsalavvllftvadvgv---nlvl---lsvlrnkklqnagnllfvslsladvav 83  

| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  

OY 61 FIFYVVNVIFIEFIEMSCANCAILLFINLEELMLATWLGVFYAKAVSVRHPLFWLMK 120  

| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  

Db 84 yypvavllalfngqwlgnlbhqsfglmgjstv-----lgsvfntatatanrycichel 138  

| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  

OY 121 RISKLK----PMHIGSLLYSMICVFSKAGFMV--PYFLKRFEQNATTIQEDTLAI 174  

| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  

Db 139 rydklyngqrstwcysiqlewtlltlatvpnfvgalsydpriifsectfaq--tvssaytltev 196

```

QY	175	QINSEFAEFSVPLPILFLFAVALLIFSLGHTNQ-----MNNTVA-----	213
Db	197	vvvhflvplsvtfcyl-ftlwtlvqvkhrvqddkqklqctldnfltmfvvlyfavc	255
OY	214	-----GSRV---PGRGAP-ISALSLISPLILVESHQMIVFLSSKPHRR	256
Db	256	waplnflglavlnphvvepkhpewlfvlsyfmayfnsclnavlygvlnqnfirk	309
RESULT	9		
ID	W25927	standard; Protein: 354 AA.	
XX	W25927;		
XX			
DT	10-NOV-1997	(first entry)	
XX			
DE	Xenopus melatonin receptor MEK-1ab.		
XX			
KM	Allele: Xenopus laevis; melatonin; receptor; untranslated region; PCR;		
KM	mRNA; half-life; skin; amplification; primer; polymerase chain reaction;		
KM	transmembrane domain; cellular signaling; inhibitor; adenylyl cyclase;		
XX	modulation; Intracellular; cyclic GMP; inhibitor; phosphodiesterase.		
XX			
OS	Xenopus laevis.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	1..32	
FT		/note= "Extracellular N-terminal domain"	
FT	Domain	33..56	
FT		/note= "transmembrane domain 1"	
FT	Domain	57..68	
FT		/note= "intracellular domain 1"	
FT	Domain	69..92	
FT		/note= "transmembrane domain 2"	
FT	Domain	93..107	
FT		/note= "extracellular domain 2"	
FT	Domain	108..125	
FT		/note= "transmembrane domain 3"	
FT	Domain	126..151	
FT		/note= "intracellular domain 2"	
FT	Domain	152..171	
FT		/note= "transmembrane domain 4"	
FT	Domain	172..193	
FT		/note= "extracellular domain 3"	
FT	Domain	194..219	
FT		/note= "transmembrane domain 5"	
FT	Domain	220..243	
FT		/note= "intracellular domain 3"	
FT	Domain	244..268	
FT		/note= "transmembrane domain 6"	
FT	Domain	269..280	
FT		/note= "extracellular domain 4"	
FT	Domain	281..301	
FT		/note= "transmembrane domain 7"	
FT	Domain	302..354	
FT		/note= "intracellular C-terminal domain"	
XX			
PN	W09704094-A1.		
XX			
PD	06-FEB-1997.		
XX			
PP	24-JUL-1996;	96MO-FR01167.	
XX			
PR	24-JUL-1995;	95FR-0008947.	
XX			
PA	(ADIR ) ADIR & CIE.		
PI	Jockers R, Marullo S, Strosberg AD;		
XX			
DR	WPI; 1997-132635/12.		
XX			
XX	N-PSDB; T79065-6.		
XX			

PT New nucleic acid encoding functional melatonin receptor of *Xenopus* -  
PT for screening for potential (ant)agonists useful for e.g. treating  
PT cardiovascular disease and cancer  
PS Claim 5; Page 30-31; 62pp; French.  
XX Sequences W25926-7 represent novel allelic *Xenopus laevis* melatonin  
CC receptor MEL-1a proteins. This sequence is a new form of the melatonin  
CC receptor designated MEL-1ab, also known as Mel 1-c(beta). The protein  
CC differs from previously known receptors by being 65 amino acids shorter  
CC and also differs throughout the sequence by 6 amino acids. The protein  
CC is encoded by 2 different alleles (779065-6) which differ in the 3',  
CC untranslated region (both sequences encode the same protein), which is  
CC thought to affect the half-life of the mRNA. The coding sequence was  
CC isolated from cDNA derived from *Xenopus* skin RNA and amplified using the  
CC primers 779067-76. The nucleotide sequence encodes a 7 transmembrane  
CC receptor involved in cellular signalling. MEL-1ab has been shown to  
CC modulate intracellular cGMP, esp. inhibiting its accumulation induced by  
CC an inhibitor of phosphodiesterase, but unlike the MEL-1a protein, it  
CC cannot inhibit adenylyl cyclase activity.  
XX  
XX Sequence 354 AA;  
SQ

Query Match 7.88; Score 118; DB 18; Length 354;  
Best Local Similarity 19.48; Pred. No. 1.2e-05;  
Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12;

OY 2 LESHLIIFLAVIOFLGFTNGIIVVNGIDILKRRMAPL-DLLSCLAVSRIFQL 60  
DB 31 lsaalavalfivadv9-----nlvli---lsvlnkklqnaqlfivsisiadvav 83  
OY 61 FIFYVNVIVIFPIEFIMCSANCALLFNELEMLWMLGVFYCAKVASVRHPLFIWLM 120  
DB 84 yypvillatfignwtclgnihcqlsgflmglsav-----lgsvfntalainrycylchsl 138  
OY 121 RISKLV----PKMILGSLIVSMICVFHSKVGFMV--PYFIRKFEFSONATQKEDTLAI 174  
DB 139 rydklfngfstfyigltwlcltlaivpufivgslyqdrfiscdaq--lvsssyctlv 196  
OY 175 GFESFAEFSVPLILFLFVALLIFSLGRTRQ-----MNTVA----- 213  
DB 197 vvvhlvlpisvvtfcyl-rivwlvigvkhrrvgddfkqrltptcdlnflmfvvlfavc 255  
OY 214 -----GSRV--PGRGAP-ISALLSILFLFYSHCMKIVPLSSLKPHIR 256  
DB 256 waplntfiglavalnplhvapkipevlfvlsymayfnsclnavlygllnqnfrk 309

RESULT 10  
R88409 standard; Protein: 420 AA.  
XX R88409;  
XX 15-AUG-1996 (first entry)  
XX High-affinity melatonin receptor.  
XX  
XX Melatonin receptor: G-protein-coupled receptor; glycosylation;  
KW disulfide bond; ligand binding pocket; phosphorylation;  
KW cytochrome c family heme binding site; melatonin receptor-agonist;  
KW melatonin receptor-antagonist; circadian rhythm disorder; jet-lag;  
KW day-night cycle disorder; ovulation; reproductive cycle; antibody;  
XX animal breeding; puberty; transgenic animal; drug screening.  
XX  
XX *Xenopus laevis*.  
OS  
XX  
XX Key Location/Qualifiers  
FH Domain 1..32  
FT /note= "Extracellular domain"  
FT Modified-site 5..7  
FT /note= "N-glycosylation site"

FT Modified-site 16  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Domain 33..57  
FT /note= "Transmembrane region-I"  
FT Domain 57..68  
FT /note= "Intracellular loop"  
FT Region 67..72  
FT /note= "Conserved melatonin receptor motif"  
FT Domain 69..93  
FT /note= "Transmembrane region-II"  
FT Domain 94..107  
FT /note= "Extracellular loop"  
FT Disulfide-bond 105..182  
FT Domain 108..126  
FT Region 125..131  
FT /note= "Region used to construct primer"  
FT Domain 127..151  
FT /note= "Intracellular loop"  
FT Region 132..137  
FT /note= "Putative cytochrome-c family heme binding site"  
FT Modified-site 137  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Domain 152..171  
FT /note= "Transmembrane region-IV"  
FT Misc-difference 166  
FT /note= "Residue which may form ligand binding pocket"  
FT Domain 172..193  
FT /note= "Extracellular loop"  
FT Domain 194..220  
FT /note= "Transmembrane region-V"  
FT Misc-difference 204  
FT /note= "Residue which may form ligand binding pocket"  
FT Domain 221..243  
FT /note= "Intracellular loop"  
FT Domain 244..268  
FT /note= "Transmembrane region-VI"  
FT Region 252..259  
FT /note= "Region used to construct primer"  
FT Misc-difference 258  
FT Domain 269..279  
FT /note= "Residue which may form ligand binding pocket"  
FT Domain 280..301  
FT /note= "Extracellular loop"  
FT Region 296..300  
FT /note= "Transmembrane region-VII"  
FT Domain 302..420  
FT /note= "Conserved melatonin receptor-1b motif"  
FT /note= "Intracellular domain"  
FT Modified-site 320  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 328  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 362  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 395  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 418  
FT /note= "Protein-kinase-C phosphorylation site"

W09535320-A1.  
28-DEC-1995.  
07-JUN-1995; 95WO-US07360.  
XX 06-JUN-1995; 95US-0319887.  
XX 17-JUN-1994; 94US-0261857.  
XX 07-OCT-1994; 94US-0319887.  
XX  
XX (MASS-) MASSACHUSETTS GEN HOSPITAL.  
XX Reprint SM;  
PI

XX MPI: 1996-058368/06.  
DR N-PSDB: T09947.  
XX  
PT DNA encoding high affinity melatonin receptor one - used to identify  
PT receptor agonists or antagonists e.g. for regulating circadian  
PT rhythm disorders or reproductive cycles  
PS  
PS Claim 5; Fig 1; 115pp; English.

CC The sequence represents a high-affinity melatonin receptor (MO).wt.  
CC 47,424) from *Xenopus laevis*. The receptor is a membrane protein,  
CC coupled to guanine nucleotide binding proteins (G-proteins), and  
CC has 7 hydrophobic putative transmembrane domains. The N-terminus  
CC has an N-linked glycosylation site typical for G-protein-coupled  
CC receptors, and 2 Cys residues in the last 2 extracellular loops may  
CC form a stabilizing disulfide bond. pro residues in transmembrane  
CC domains IV, V and VI may introduce kinks in the alpha-helices to  
CC form of a ligand binding pocket. Phosphorylation sites in the  
CC C-tail may be involved in receptor regulation. Primers from the  
CC encoding DNA may be used for isolation of sheep, mouse and human  
CC receptor sequences. Receptor fragments which interact with  
CC melatonin, or specific antibodies, may be used as receptor-agonists  
CC or receptor-antagonists. Agonists may be used in therapy of  
CC circadian rhythm disorders such as jet-lag or day-night cycle  
CC disorders, to control ovulation, or in alteration of reproductive  
CC cycles in seasonally breeding animals. Antagonists may be used to  
CC control the initiation or timing of puberty in humans. The  
CC receptor gene may also be expressed in a transgenic animal for use  
CC as a model system to screen agonists and antagonists.

XX Sequence 420 AA:

Query Match 7.8%; Score 118; DB 17; Length 420;  
Best Local Similarity 19.4%; Pred. No. 1.5e-05;  
Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12;

OY 2 LESHLITFLAVAYQPLIGFTNGIIVNGIDILKRRKAPL-DLLSCLAVSRIFQL 60  
DB 31 ltsalavvllflvadvlg---nlv1---lavlnkkikqngnlfvvalsladvlavv 83  
OY 61 FIFVYVIVFEFEFICSNCAILFELNELMLATWIGVFCANVASRHHFLFMK 120  
DB 84 YPPVVIILAFIqgwltgnlncqsgflmglyv---lgsvtntalainrycchsl 138  
OY 121 RISKLV---PMWILGSLVSMICVFHSKYAGFMV--PYELRRFSONATIQEDTLAI 174  
DB 139 rydklylqrstcwyglitwlltllataypntffvgslyqdpriacteq--tvesstlylv 196  
OY 175 QIFSFAEFSVPPLILFIFAFVLLIFSLGRTRQ-----MRNTVA----- 213  
DB 197 vvvhflvplsvlfcyl-rltwlvlvqvkhrvqdfkqklctqdlrflmfvfvlfavv 255  
OY 214 -----GSRV---PGRGAP-ISALLSILSLFLYFSHCKIKVFLSLKFRHNR 256  
DB 256 waplntflglavlnpflvapkdpewlfvlsyfmayfnsclnavlyvlnqnfik 309

RESULT 11  
W19105 standard; Protein; 310 AA.  
XX W19105;  
XX  
XX 26-AUG-1997 (first entry)  
XX  
XX Rat pheromone receptor VN4.  
DE  
KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;  
KW maternal behaviour; reproductive behaviour; fertility;  
KW hormone secretion.  
XX

OS Rattus sp.  
XX  
XX Key Location/Qualifiers  
FH Domain 16..37  
FT /label= I  
FT /note= "predicted position of transmembrane domain  
FT I"  
FT Domain 50..71  
FT /label= II  
FT /note= "predicted position of transmembrane domain  
FT II"  
FT Domain 97..116  
FT /label= III  
FT /note= "predicted position of transmembrane domain  
FT III"  
FT Domain 135..156  
FT /label= IV  
FT /note= "predicted position of transmembrane domain  
FT IV"  
FT Domain 194..212  
FT /label= V  
FT /note= "predicted position of transmembrane domain  
FT V"  
FT Domain 240..258  
FT /label= VI  
FT /note= "predicted position of transmembrane domain  
FT VI"  
FT Domain 270..292  
FT /label= VII  
FT /note= "predicted position of transmembrane domain  
FT VII"  
FT Misc-difference 261  
FT /note= "Ser-261 is additional to the amino acid  
FT residues deduced from the nucleotide  
FT sequence"  
FT Misc-difference 263..264  
FT /note= "the amino acid sequence deduced from the  
FT nucleotide sequence has a lysine residue  
FT between Met-263 and Phe-264"  
PN W09714790-A1.  
XX  
XX 24-APR-1997.  
PD  
XX 18-OCT-1996; 96MO-US16637.  
PF  
XX 19-OCT-1995; 95US-0005698.  
PR  
XX (UYCO ) UNITV COLUMBIA NEW YORK.  
PA  
XX Axel R, Dulac C;  
PI  
XX MPI: 1997-245107/22.  
DR N-PSDB: T69547.  
DR  
XX  
PT Nucleic acid molecule encoding vertebrate pheromone receptor -  
PT useful to identify modulators for control of reproductive and social  
PT behaviour, fertility and hormone secretion  
PS  
PS Disclosure: Page 79-80; 123pp; English.

CC Rat pheromone receptors VN1, VN3, VN4, VN5, VN6, VN7 and VN2  
CC (W19103-09) are members of a novel family of presumed  
CC 7-transmembrane domain receptors that are evolutionary independent  
CC of the odorant receptors of the main olfactory epithelium (MOE).  
CC Their amino acid sequences were deduced from cDNA clones (see also  
CC T69545-50. VN2 sequence not provided) obtd. by differential  
CC cloning, PCR and hybridisation from single vomeronasal sensory  
CC neurons (VSN). Individual neurons express different complements of  
CC the receptors. VN polypeptides can be expressed in host-vector  
CC systems for use in identifying modulators for control of maternal,  
CC reproductive and social behavior, to increase fertility, control  
CC hormone secretion and to regulate food uptake in humans and



PR 18-SEP-1997: 97US-0059263.  
 PR 18-SEP-1997: 97US-0059266.  
 PR 15-OCT-1997: 97US-0062155.  
 CC 17-OCT-1997: 97US-0062285.  
 PR 17-OCT-1997: 97US-0062287.  
 PR 21-OCT-1997: 97US-0063486.  
 PR 24-OCT-1997: 97US-0062814.  
 PR 24-OCT-1997: 97US-0062816.  
 PR 24-OCT-1997: 97US-0063045.  
 PR 24-OCT-1997: 97US-0063120.  
 PR 24-OCT-1997: 97US-0063121.  
 PR 24-OCT-1997: 97US-0063127.  
 PR 24-OCT-1997: 97US-0063128.  
 PR 27-OCT-1997: 97US-0063329.  
 PR 27-OCT-1997: 97US-0063327.  
 PR 28-OCT-1997: 97US-0063541.  
 PR 28-OCT-1997: 97US-0063542.  
 PR 28-OCT-1997: 97US-0063544.  
 PR 28-OCT-1997: 97US-0063549.  
 PR 28-OCT-1997: 97US-0063550.  
 PR 28-OCT-1997: 97US-0063564.  
 PR 29-OCT-1997: 97US-0063435.  
 PR 29-OCT-1997: 97US-0063704.  
 PR 29-OCT-1997: 97US-0063732.  
 PR 29-OCT-1997: 97US-0063738.  
 PR 29-OCT-1997: 97US-0063734.  
 PR 29-OCT-1997: 97US-0064215.  
 PR 29-OCT-1997: 97US-0063735.  
 PR 31-OCT-1997: 97US-0063870.  
 PR 31-OCT-1997: 97US-0064103.  
 PR 03-NOV-1997: 97US-0064248.  
 PR 07-NOV-1997: 97US-0064809.  
 PR 12-NOV-1997: 97US-0065186.  
 PR 17-NOV-1997: 97US-0065846.  
 PR 18-NOV-1997: 97US-0065693.  
 PR 21-NOV-1997: 97US-0066120.  
 PR 21-NOV-1997: 97US-0066364.  
 PR 24-NOV-1997: 97US-0066772.  
 PR 24-NOV-1997: 97US-0066466.  
 PR 24-NOV-1997: 97US-0066770.  
 PR 24-NOV-1997: 97US-0066511.  
 PR 24-NOV-1997: 97US-0066453.  
 XX (GETH ) GENENTECH INC.  
 PA Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
 PI WPI: 1999-229533/19.  
 DR N-PSDB; X52222.  
 XX  
 PT New isolated human genes and polypeptides used in, e.g. treatment of  
 PT gastrointestinal ulceration  
 XX  
 PS Claim 12; Fig 19; 320pp; English.  
 XX  
 CC Y13344-003 represent secreted and transmembrane human proteins.  
 CC The cDNA sequences are obtained from cDNA libraries, prepared from  
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.  
 CC The encoded polypeptides have specific uses based on their homology to  
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders  
 CC associated with the preservation and maintenance of gastrointestinal  
 CC mucosa and the repair of acute and chronic mucosal lesions  
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal  
 CC ulceration and congenital microvillus atrophy), skin diseases associated  
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
 CC potent effects on cell growth and development, diseases related to growth  
 CC or survival of nerve cells including Parkinson's disease, Alzheimer's  
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for  
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment  
 CC of Usher Syndrome or Atrophid areata; PRO269 can be used as an  
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have

CC therapeutic applications in wound healing and tissue repair; PRO317 can  
 CC be used for treating problems of the kidney, uterus, endometrium, blood  
 CC vessels, or related tissue, e.g. in the heart of genital tract.  
 XX  
 SQ Sequence 690 AA:  
 Query Match 7.7%; Score 117.5; DB 20; Length 690;  
 Best Local Similarity 21.3%; Pred. No. 3.3e-05;  
 Matches 71; Conservative 59; Mismatches 114; Indels 89; Gaps 16;  
 OY 4 SHLITVELAVIQFLGIFTNGIIVVNGIDLKHKMAPDLDSCLAVSRFLDLTF 63  
 DB 404 nhltfaiimsqpsigldkymllrtltqgil-----islclalc-----lftc 449  
 OY 64 YVNVIVFFIEF-----IMCSANCAILFELNELMLATWLGVEYCAKVASVRHP 113  
 DB 450 w-----ffaelqstrtlclhkcslflaelvfl-----vgintntnklfcslagllhy 499  
 OY 114 LFIWLMKMRISKLVPMW-IIGSLLYVSMICVPHSKVGEW-VYFPLRKFSONATTOKEDT 171  
 DB 500 ffl-----aafamwleglhllylvvglynk--gflhknfyllfgyispavvvvffsaa 550  
 OY 172 LAIQ-----IFSVAEFSVPLIFPLFAVLILFISGRTRQMRNTVA-- 213  
 DB 551 lgyryygttkvcwslstemffwafgpacllllvnlafgvllykvrhtaglkpevscl 610  
 OY 214 -GSRVGRGAPISALISITIS---FLILYFSHOMI---KVPLSSLKPHIRPIFFFLIV 265  
 DB 611 enlrsarga--lallflgtwtlfgvlhvnavsavtlaylfvsnafq-gmfelflfcvl 667  
 OY 266 IGYPGSHSLIILGNPKLONAKKFLHSCC 298  
 DB 668 -----srklgeeyylfknpcc 685  
 RESULT 14  
 Y83225  
 ID Y83225 standard; Protein; 690 AA.  
 XX  
 AC Y83225;  
 XX  
 DT 16-AUG-2000 (first entry)  
 XX  
 DE PRO288 Polypeptide.  
 XX  
 KW Inhibition: cancer; neoplasia; tumour; breast; ovary; renal;  
 KW colorectal; uterus; prostate; lung; bladder; central nervous system;  
 KW CNS; melanoma; leukaemia; PRO211; PRO228; PRO318; PRO182;  
 KW human; ss  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FH Peptide  
 FT 1.19 Location/Qualifiers  
 FT 15..19 /label= Signal\_peptide  
 FT /note= "N-glycosylation site"  
 FT 21..25 /note= "N-glycosylation site"  
 FT /note= "N-glycosylation site"  
 FT 38..44 /note= "N-myristoylation site"  
 FT /note= "N-myristoylation site"  
 FT 50..56 /note= "N-myristoylation site"  
 FT 4953..53 /note= "Glycosaminoglycan attachment site"  
 FT 52..58 /note= "N-myristoylation site"  
 FT /note= "N-myristoylation site"  
 FT 54..58 /note= "N-myristoylation site"  
 FT /note= "Casein kinase II phosphorylation site"  
 FT 64..68 /note= "N-glycosylation site"  
 FT Modified-site  
 FT Modified-site  
 FT Modified-site



